

STIC-Biotech/ChemLib

8B13

8E12

From: Davis, Natalie
Sent: Tuesday, August 27, 2002 5:08 PM
To: STIC-Biotech/ChemLib
Subj ct: 09/807470

Please search SEQ ID NO: 1-4 for 09/807470 and interference.

Natalie A.Davis, PhD
Patent Examiner
Art Unit 1642
CM1, Rm 8B13
Mailbox 8E12
Ph (703) 308-6410

Point of Contact
Susan Hanley
Technical Info. Specialist
CM1 6B05 Tel: 305-4053

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

Point of Contact
Susan H. Hines
Technical Info Specialist
CMI 8800 Tel. 303-4053

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 04:39:56 ; Search time 5654.61 Seconds
(without alignments)
8552.531 Million cell updates/sec

Title: US-09-807-470-1

Perfect score: 2311

Sequence: 1 agagagagagagagagag.....aaaaaaaaaaaaaaaaaaaaa 2311

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score Match	Length	DB ID	Description
------------	-------------	-------------	--------	-------	-------------

RESULT 1

AC106086

LOCUS

DEFINITION

AC106086

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

ALIGNMENTS

AC106086

LOCUS

DEFINITION

AC106086

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC106086 Rattus no
BD005227 Method of
AC093805 Homo sapi
BD005228 Method of
BD005226 Method of
X63678 C.familiari
AC000687 Homo sapi
AC329795 Sequence
AC005229 Method of
X63679 H.sapiens m
BC012401 Mus muscu
U19578 Bos taurus
AY029764 Mus muscu
AY029534 Xenopus l
AY029529 Danio rer
AX303127 Sequence
D31762 Human mRNA
BC018212 Mus muscu
AY029530 Mus muscu
AY029535 Xenopus l
AY075393 Drosophil
AC014319 Drosophil
AC097726 Drosophil
AE003419 Drosophil
AL109630 Drosophil
AC027246 Homo sapi
AC022731 Homo sapi
AC022032 Homo sapi
AU026177 Rattus no
AC107646 Mus muscu
AC022032 Homo sapi
AL049382 Homo sapi
AC110042 Mus muscu
AB070034 Macaca fa
AC068141 Mus muscu
AX344566 Sequence
Z67988 C.chinensis
AL390756 Homo sapi
BC018189 Homo sapi
AX092338 Sequence
BC017614 Mus muscu
AX344936 Sequence
AC084812 Homo sapi
AX185705 Sequence
L29571 Aequorea vi

linear

HTG 12-JAN-2002

SEQUENCING IN PROGRESS

AC106086

Rattus norvegicus clone CH230-126G18, ***

15623 bp DNA

***, 70 unordered pieces.

AC106086

AC106086.1 GI:18138601

HTG; HTGS_PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 15623)

Muzny,D.M., Adams,C., Ali-Osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbaria,J.,

Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carroon,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

```

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,D.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisgeed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzkler,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaite,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczky,K., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 155623)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHRR
Center clone name: CH230-126G18
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 127470 bases at least Q40
Consensus quality: 134419 bases at least Q30
Consensus quality: 139475 bases at least Q20
Estimated insert size: 122947; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-tp estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5824: contig of 5824 bp in length
* 5825 5924: gap of unknown length
* 5925 10293: contig of 4369 bp in length
* 10294 10393: gap of unknown length
* 10394 13928: contig of 3535 bp in length
* 13929 14028: gap of unknown length
* 14029 19438: contig of 5410 bp in length
*
19439 19538: gap of unknown length
19539 22725: contig of 3187 bp in length
22726 22825: gap of unknown length
22826 24943: contig of 2118 bp in length
24944 25043: gap of unknown length
25044 29216: contig of 4173 bp in length
29217 33015: contig of 3699 bp in length
33016 33115: gap of unknown length
33116 37717: contig of 4602 bp in length
37718 40275: contig of 2458 bp in length
40276 40375: gap of unknown length
40376 42400: contig of 2025 bp in length
42401 42500: gap of unknown length
42501 45123: contig of 2523 bp in length
45124 48116: contig of 2993 bp in length
48117 48216: gap of unknown length
48217 51729: contig of 3513 bp in length
51730 51829: gap of unknown length
51830 54589: contig of 2760 bp in length
54590 54690: gap of unknown length
54691 57532: contig of 2843 bp in length
57533 60697: contig of 3065 bp in length
60698 60797: gap of unknown length
60798 62618: contig of 1821 bp in length
62619 62718: gap of unknown length
62719 65152: contig of 2433 bp in length
65153 67843: contig of 2592 bp in length
67844 67943: gap of unknown length
67944 69284: contig of 1321 bp in length
69285 71847: contig of 2483 bp in length
71848 74812: contig of 2865 bp in length
74813 74912: gap of unknown length
74913 77983: contig of 3071 bp in length
77984 78083: gap of unknown length
80155 80255: contig of 2072 bp in length
80256 81813: contig of 1558 bp in length
81814 83742: contig of 1829 bp in length
83743 83842: gap of unknown length
83843 86103: contig of 2261 bp in length
86104 88411: contig of 2208 bp in length
88412 90397: contig of 1886 bp in length
90398 90497: gap of unknown length
90498 92759: contig of 2262 bp in length
92760 94270: contig of 1411 bp in length
94271 96506: contig of 2136 bp in length
96507 96606: gap of unknown length
96607 97996: contig of 1390 bp in length
97997 98096: gap of unknown length
98097 99954: contig of 1858 bp in length
99955 100054: gap of unknown length
100055 101350: contig of 1296 bp in length
101351 101450: gap of unknown length
101451 103336: contig of 1886 bp in length
103337 104996: contig of 1560 bp in length
104997 105096: gap of unknown length
105097 107126: contig of 2030 bp in length
107127 107226: gap of unknown length
107227 108688: contig of 1442 bp in length
108689 108768: gap of unknown length

```

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

QY 1363 tcgagttgcagttccaggtgtacataacatgacottgacacctgctggttcagaga 1422
 Db 901 TCAGTTGCAGTATCCAGGTGTACATAACATGGACCTTGACCGCTCGCTTCAGAGA 960
 QY 1423 tggtagaagatcgcaatcttcgtctgtggaggaagagacggtccaggtcgagaaaa 1482
 Db 961 TGCTTAGAAGATCGCAATCTTCATGTCTGTGGGGAAGAGACGGTCCAGTTCGAGAAA 1020
 QY 1483 ggcacagaaatggagtggagaatccaaatagatattctccaccacaaagaagaagag 1542
 Db 1021 GGCACAGAAATGGAGTGGAGAAATCCAAATAGATATTCTCCACCAAGAAAGAG 1080
 QY 1543 aaagctccttag 1554
 Db 1081 AAAGCTCCTTAG 1092

RESULT 3
 AC093805/c
 LOCUS Homo sapiens BAC clone RP11-32609 from 4, complete sequence.
 DEFINITION AC093805 AC012532
 ACCESSION AC093805.3 GI:18497223
 VERSION HTG.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 JOURNAL 1 (bases 1 to 165538)
 MEDLINE Sulton, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 REFERENCE 99083792
 2 (bases 1 to 165538)
 Cedroni, M., Abbott, A. and Bielicki, L.
 TITLE The sequence of Homo sapiens BAC clone RP11-32609
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 165538)
 Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (10-SEP-2001) Genome Sequencing Center, Washington
 University School of Medicine, 444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 4 (bases 1 to 165538)
 Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (05-FEB-2002) Genome Sequencing Center, Washington
 University School of Medicine, 444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 5 (bases 1 to 165538)
 Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (08-FEB-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Feb 5, 2002 this sequence version replaced gi:16077043.
 COMMENT ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0326009
 Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.
 This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-32609; actual end is at base position 165538 of RP11-32609.

Data from AC013556 was used to finish the clone, AC093805. Polymorphisms have been identified between AC013556 and AC093805.

The sequence of AC012532 has been incorporated into AC093805.

FEATURES

source	Location/Qualifiers
1. .165538	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-32609"
	/clone_lib="RPCI-11"
	/rpt_family="MaLR"
	131. .436
	/rpt_family="ERV1"
	870. .1245
	/rpt_family="L1"
	1246. .2720
	/rpt_family="L1"
	2721. .3067
	/rpt_family="L1"
	3627. .3713
	/rpt_family="MaLR"
	4217. .5074
	/rpt_family="AchoBo"
	5055. .5101
	/rpt_family="(TTTA)n"
	5075. .5384
	/rpt_family="Alu"
	5385. .5592
	/rpt_family="AchoBo"
	7152. .7426
	/rpt_family="L1"
	7427. .7734
	/rpt_family="Alu"
	7590. .7619
	/rpt_family="AT-rich"
	7735. .7872
	/rpt_family="L1"
	7873. .8314
	/rpt_family="ERV1"
	8315. .9564
	/rpt_family="L1"
	9438. .9470
	/rpt_family="(A)n"

repeat_region	9822.._9853
	/rpt_family="(CA)n"
repeat_region	10374.._10722
	/rpt_family="MaLR"
repeat_region	11045.._11163
	/rpt_family="CT-rich"
repeat_region	13136.._13287
	/rpt_family="Alu"
repeat_region	15545.._15567
	/rpt_family="AT-rich"
repeat_region	15850.._15875
	/rpt_family="AT-rich"
repeat_region	15929.._16234
	/rpt_family="Alu"
repeat_region	17161.._17194
	/rpt_family="AT-rich"
repeat_region	18540.._18828
	/rpt_family="MaLR"
repeat_region	18938.._19023
	/rpt_family="MaLR"
repeat_region	19181.._19224
	/rpt_family="MaLR"
repeat_region	19256.._19598
	/rpt_family="MaLR"
repeat_region	19659.._19726
	/rpt_family="MaLR"
repeat_region	19847.._20312
	/rpt_family="MER2_type"
repeat_region	19928.._19960
	/rpt_family="(TG)n"
repeat_region	20313.._20623
	/rpt_family="Alu"
repeat_region	20596.._20640
	/rpt_family="(A)n"
repeat_region	20779.._20802
	/rpt_family="AT-rich"
repeat_region	20868.._21156
	/rpt_family="Alu"
repeat_region	21487.._21514
	/rpt_family="AT-rich"
repeat_region	21958.._21983
	/rpt_family="AT-rich"
repeat_region	22010.._22076
	/rpt_family="AT-rich"
repeat_region	22223.._22248
	/rpt_family="(TTTA)n"
repeat_region	22234.._22520
	/rpt_family="Alu"
repeat_region	23084.._23385
	/rpt_family="Alu"
repeat_region	23503.._23691
	/rpt_family="CR1"
repeat_region	23715.._23895
	/rpt_family="MIR"
repeat_region	24054.._24855
	/rpt_family="CR1"
repeat_region	24966.._25190
	/rpt_family="L2"
repeat_region	25636.._25670
	/rpt_family="(CA)n"
repeat_region	25704.._26073
	/rpt_family="MaLR"
repeat_region	26711.._27112
	/rpt_family="MaLR"
repeat_region	27461.._27664
	/rpt_family="L2"
repeat_region	28756.._28814
	/rpt_family="L2"
repeat_region	29356.._29574
	/rpt_family="L1"
repeat_region	29748.._29876
	/rpt_family="ERV1"
repeat_region	30292.._30476

```

/rptc_family="L1"
Query Match      35.0%; Score 808.4; DB 9; Length 165538;
Best Local Similarity 67.0%; Pred. No. 6.9e-152;
Matches 1478; Conservative 0; Mismatches 621; Indels 107; Gaps 19;

QY 194 cggctcccgcggtggcgcgcgcgatcggtgctggcgaatacgcctcgacccccggg 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 49264 CGGTGGCAGCGAGTTTGAAGATGCGACGATCAACGTTGAAGATCACCGCTCGCAACCCG 49205

QY 254 -----cctgcgggcgagggcgcgcgctcgatcttcctcctcctctcgag---cccc 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 49204 GGCCTGGCGCGGGTAGGGGCGCGCGCTCGATTTCTTCCTCGCTCCGCGCTCCCT 49145

QY 306 tgtgcgcatgctcggc-ctacgcgcccccgaccttgattga-----tcgctcgcgacgc 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 49144 GGTGGCGCATGCTCAGCTCAGCTCGGCCCTCGCCTTTGATTTATTTTCTGGCGGCC 49085

QY 360 gctgcgacctggcgcgacgagcgggcggggagcg-----cggcgctggg 409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 49084 GCTGGCAGCCGGGACTGACTTCGGGATGGGAAGTGGAGCCCGCGAGCTGACTCGGTGC 49025

QY 410 agcgcgcgagtgtacgc-----ggtcggcgcggtgagtaccggtgagtaccgcgg 461
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 49024 GSGCGCGCTGTGAGGAGCAGCGACGGGGGAGCAGCTGCGGCTCGCGGTGAGTATCGGG 48965

QY 462 -----catggggtccgcaagaagaaacgcagagaaccccccggtgtgtgaccacga 512
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48964 AAGCGCCACCATTGGGGCTCCGTAAGAAGAGACCAAGAACCCCGCTTCAGCCAGGA 48905

QY 513 attcatggtgcagaacacgcgatatggtctcctcgctgggcgatgtcttcgtgctggg 572
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48904 ATTCACTCGCAGAAATCATGCGGACATCGTCTCTCGGTGGGGATGTTCTTCTGTCTGG 48845

QY 573 acttatgttcagggcacggccgagatgtcgatgtcttctcaacctgcagcatggagt 632
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48844 GCTTGTGTTCAGGGGAACAGCAGAGAAGCATCATCGTGTCTCTCACTTTCAGCACAGTGT 48785

QY 633 cgtgtcc---cagcggaagggctaccctcgggttcagagaccctttaccatattgggt 689
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48784 TGCTGTCCCTGCACACAGAGNACAAAGCCAGGGCTCAAAGTCCTCTATTATTATGTGTGT 48725

QY 690 caaagatctggccacagtgttcttctacatgctggtggccatcatcatcaccgccacat 749
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48724 CAAAGATTGGCCACGGTTTCTTCTACATGCTGGTGGCAATCATTTATTCATGCCAAT 48665

QY 750 tcaggagtacgtctaatgaactcaacgcgagactgcagctcaccaaaggcacaacaaa 809
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48664 TCAGGAATATGTGTGGATAAATTAACAAGAGAATGCATTCACCAAGCGAAGCAAA 48605

QY 810 caaatgaattagcgccggcgagctgagtgttctcacaatgctctcgttatctggggtat 869
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48604 CAAAGTTTAAAGAGTCTGGTCAGTTTAGTGTGTCTACTTTTTTCTGTATTATTTGGGCAC 48545

QY 870 gatcattcggctctgaagaactgcctgtcagaccccccaactctattgtgaagctcagcc 929
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48544 ATTCAATTTAATCTCTGAAACTGCGGTGTCAGACCAACTCTTTATATGAAAGGCTCGTCC 48485

QY 930 ccacaacatgatgacatttcagatgaattttctacatctcagcttgagcttaactggtt 989
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48484 CCATAGCATGATGACATTTCAATGAAGTTTTCACATATCCAGTTGGCTTACTTGGTT 48425

QY 990 tcatagttcccgagctctacttcagaaaagtccaggaacaaagatatccccgggtcaact 1049
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48424 TCATGCTTTTTCGTAACCTCTACTTCAGAAAACCAAAAACAGACATCCCTCGTCAACT 48365

QY 1050 catctacattggctccacacctctcacattggaggggcctatctctctgacttgaaacca 1109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48364 TGCTACATTTGGCTTCACCTCTCCACATTACTGGAGCTTATCTCTGTACTTGATCA 48305

QY 1110 cctgggctcgtcctctgatgctgcactatgctgcagagctcctccagcaggtgtgcag 1169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48304 TTTGGGACCTCTCTTTGGTATCTGATATTTTGTGTAATTAATTAATTTCCACATATGCGG 48245

```

```
QY 1170 cctgcttactttgggatgagcgggtaccagaaggggttgctcttcttgccctatcgtgt 1229
DB 48244 CCGTGTCTTACTTTAGTGATGAAAAGTACCAGAAAAGCATATCTCTGTGGCCCATGTGT 48185
QY 1230 tatatccgggagactcgtgacactgattgtctcagtggttacaagttaggggttcacttggc 1289
DB 48184 TATCTGGGTAGACITGTGCATTTAATTGTTCCGTFACCTACCTGTTGGGTTTCACCTGGC 48125
QY 1290 cgg---gacaaatcggaatcggaatgctctctctgtgtaatgtcaatgtgttggcagctaa 1346
DB 48124 TGGATGCGCAAGATCGGAATCTCGATCCCTTACTGGAATGTAAATGTGTGGCAGCTAA 48065
QY 1347 aatcgcgtcttctgctcagttcagttccagtggttacaatgagttcagttcagttcagttc 1406
DB 48064 AATTCGTGTCTGTCTCCAGTTGTCAGATCCAGCTCAAGCTACGTAACTGAACCTTAATTAC 48005
QY 1407 cgtctggcttcagagatggttagaagatcggaatcttctatgctgtggaggaa----- 1460
DB 48004 TCTCTGGCTTCAGAGTGGGTAGAGATCTAATAATTCAGGCCCTCATGTATGAAGAA 47945
QY 1461 -----gagacggtccagtcagagaaagccacagaaaatggagtggtgagaatccaaatag 1514
DB 47944 ACGGTGCGAGATCTCTAAAAAAGAAAGACAGAAAACGGAGTGGGAGTGGAAACTTCAAAATAG 47885
QY 1515 aatagaattctccacaaagaaagaaagaaagcctcttagcattgcaagcaattgatt 1574
DB 47884 AGTAGACTCTCCCAAGAGAGAAAGAAATCTTCAATACTTGCAGCGCATTTGATT 47825
QY 1575 cttacatccaaagggaatccactctctctatgtggtgtctctgtctagaga-ttttctg 1633
DB 47824 AATGTCTGCAAAAGGAAT---CTGCTCTTTGAGGTTTCTTTCTGCACTAGAGATTTTCTG 47768
QY 1634 tcttcagaacgggtcgctgttttgaatattgctaatgtattgtcttaattgttctt--- 1691
DB 47767 TTTTGTAAAAATAGTCTGCTCTCTCGGTTTGTGTTATTGAACCTGTTTCATGTATTTTGA 47708
QY 1692 -----aaggttttcagacgctatgagtggggagtggggtttaaagacta 1734
DB 47707 AAGACATTTTGGGGGAGGAGGATTATATGAATGGGAAAAAAGATTTTGGTTGACACTA 47648
QY 1735 aaccactcagcctctaatacagtcagagatagttaaacg---accacatcttatttag 1790
DB 47647 AATTACTACTCGTCAAAAATAATGTCAAAATAGTTTGGGGATCACCACTATATTTTGTGT 47588
QY 1791 tttaggttctacctaacaagatttccaacggttttgggtgagtcagtcagaattgtgta 1850
DB 47587 TGATTTTAACTTTCACATTTTCCCTAATGATTTGCAGAGATAAAGTGCACAAATTTTGA 47528
QY 1851 cataaataatagt-----ttcctgcttccaatgttcttcttatacgaattaaacaagtctgcta 1905
DB 47527 TATCAATGATACATGGTTCCTTACTCCACCACTGTTTCAATAAATACTAACAAGATGGTCTCT 47468
QY 1906 gaaagtgtgttttctcaatgttctcctcagagataaagtggaaatctgataaagg 1965
DB 47467 CCTAGCAAGATATGTGTTTAAATGCTTGTGTTGGGGTAAAA--TAAAGTACCAAAAAGG 47410
QY 1966 ttaaacctcaaatcaagtattatcttaacccgttgggattttttaaagtgttttaaattaca 2025
DB 47409 TGGAAAGTCAAAATCAGTATCTGTGAATTTGTAGAAATTTATTTTAAAGAACTTCAAACTCA 47350
QY 2026 atggaaagcatttgttcaaacaccaca-aaaatatgtgtttaaattttatgagtagtaattgt 2084
DB 47349 GAAAAGATTTGCTAGACTACCAAAAATAAATAATGTTCTTTATTTTACAGTAGTGAATAT 47290
QY 2085 tagtgtttagcgtcccatataagcatcaaaatgatgaatagatgacatgtgtgtgtatatt 2144
DB 47289 TAGTGTCTTATCCCAATTTAAAAAACAACACAGACTACTAATGGGTAAACACATATGGAGGTTT- 47229
QY 2145 gacatttagcgaatcaagataccttttaataaatatggtgtggttactaaagaagtaaacga 2204
DB 47230 -----GCTGCCATATATATTGTCATCAAAATATCATTAATTAATAAATAATTAATA 47179
```

```
QY 2205 cttctcctgtttattttaaacactgtacaggaataactcgcaaaataataattactactga 2264
DB 47178 ATCATTCCTGTCCATT---CCACTTGAATGGGAATTCATGAATTAATAATATTACTTTT 47123
QY 2265 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2310
DB 47122 AAAAAGGACATTGAACCTAGTCTTTTAAAAAATGTTTATTAAAAA 47077

RESULT 4
BD005228
LOCUS BD005228 1288 bp DNA linear PAT 31-JAN-2002
DEFINITION Method of screening of protein.
ACCESSION BD005228
VERSION BD005228.1 GI:18633189
KEYWORDS JP 03075332-T/3.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1288)
Todo,N., Okuyama,H., Imamura,M., Ishikawa,H. and Nemoto,K.
Method of screening of protein
Patent: JP 03075332-T 3 16-FEB-2001;
SUMITOMO PHARMACEUTICALS CO LTD,NAOKI TODO,HAJIME OKUYAMA,OTOAKI
IMAMURA,HIROKORI ISHIKAWA,KIYOMITSU NEMOTO
OS Homo sapiens (human)
PN JP 03075332-T/3
PD 16-FEB-2001
PF 17-AUG-2000 JP 2000005488
PI 20-AUG-1999 JP 99p 234764
PI NAOKI TODO,HAJIME OKUYAMA,OTOAKI IMAMURA,HIROKORI ISHIKAWA,
PI KIYOMITSU NEMOTO
PC C12Q1/02,60IN33/50,C07K14/47,A61K38/17,C12N5/10,C12P21/02// PC
CC (C12P21/02,C12RI:91)

FEATURES
source
key CDS Location/Qualifiers
FT CDS 331 a 275 c 276 g 406 t
BASE COUNT 331 a 275 c 276 g 406 t
ORIGIN

Query Match 30.1%; Score 695.2; DB 6; Length 1288;
Best Local Similarity 74.9%; Pred. No. 1.3e-129;
Matches 957; Conservative 0; Mismatches 298; Indels 23; Gaps 6;

QY 435 gccggtgagtacccgttgagtaccgcggtacgggtccgcgaagaagacgccaggaaccc 494
DB 13 GCCCGTGAGTATCCGGGA-AGCGCCACCATGGGCTCCGTAAGAAGAGACACCAAGACC 71
QY 495 ccggtgctcagccacgaattcattggtgcagaaccacgcggtatgctcctcgtggtg 554
DB 72 CCCCGTTCTCAGCCAGGAATTCCTTCGAGAATCATCGGACATCGTCTCCTCGGTGGG 131
QY 555 catgttctcgtgctgggacttatgttcgagggcagccgagatgctgactgttctct 614
DB 132 GATGTTCTTCTCGTGGGCTTGTTCGAGGGAACAGCAGCAAGCATCCATCGTGTCT 191
QY 615 caccctgcagtcagtgagtcgtgttgc---cagcgaaggggtaccctcgggtccaggac 671
DB 192 CACTCTTCAGCACAGTGTGCTGTCCCTGTCAGCAGAGAGAAAGCCACGGCTCAAAATC 251
QY 672 cctttaccattatgggtcaagatctggccacagttcttctctacatgctgggtggccat 731
DB 252 CCTCTATTATTATGTTGCAAAAGATTTGGCCACCGGTTTCTTCTACATGCTGGTGGCAAT 311
QY 732 catcattcaagccaccattccaggtacgtgtagataagctcagccggagactgcagct 791
DB 312 CATTTATTCATGCCACATTCAGGAATATGTTGTTGATAAAATTAACAAGAGANTGCATT 371
```



```
QY 1330 aatgtgttgagcagtaaaatcgtctgtctcgtcaggtgcagttacagtggtacata 1389
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1092 AATGTTGTTAGCTGTAGAATGCTGTCCTGGCATCCATTTGCATTACCAAGCATTCATG 1151
QY 1390 acatgagaccttgacgacctcgtggcttcagagatggttagaagatcgcaatcttcgtc 1449
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1152 ATGTGGAAGTTCATTAATTTTCAGCTTCGGAGGTGGAGGAACATTTCTACTTTTCAGGCA 1211
QY 1450 tttgggaggaag-----agacggtccaggtcgagaaagacagaa 1491
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1212 CCAGTGTGCAAGAAACACACAGTAACAGGAGGTCTTTCTAGAAAGGAACAGAA 1271
QY 1492 aatggagt-----ggagaatccaatagatattctccaccagaagaagaag 1542
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1272 AATGGTGTAAATGGAACAGTAACCTCAATGGAGCTGACTCCCGCTAATAGAAAGAG 1331
QY 1543 aaagc-tccttagcagtgcaagcgaatgattctctacccccaaggaatccactcttc 1601
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1332 AAATCTTCATAATGAATTAATCTAATTAATGAATTAATGTCCTCCCAAGAAATCTGCTTTTA 1391
QY 1602 ttatgggtgctcgtcgtcagagatttctctcttcaga--acgggtcggtcttttg 1659
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1392 CTATAT-CTTTCAGACTAGAGATTTTCTGTCCTTGAAATACAGTTTGTGCTCTTTG 1450
QY 1660 aatattgctaattgtcttaattgtttttaaaggttttcagacagtgatgagtgggga 1719
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1451 ATTTCTTGCTATTGCTACTGTTGTTTCATGCATT--TTTTTAAAGGCGATTTGAGGGAAGGA 1508
QY 1720 tgggggttaagactaaacactcagcct-----ctaaatacagtcagataag 1766
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1509 TGATTTACTATGATGTGAGAAAAATATTTTAGCTTAGACTAAGCTTACCTGCTTCAAAATAG 1568
QY 1767 ttaacgagcaaacatcttatttagtgctctt--acctcaacgattttcccaacgcttt 1824
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1569 TTTAGGACCAACACCATATCTATTTGTTTATCTTTTATCTTTGAATGTTTCTTAAGATT 1628
QY 1825 tgtgtgtgact-----gcagaattgtgtacataaataatagtttctcgtctccaatgt 1879
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1629 TGGAGAGAACTATTTTACAAAAATCTACATAACAGTGATAAAAATTTCTTGCTCACC 1688
QY 1880 tctttatgaataaacaag-----tctgtacgaagtggtttgttttctcgaattct 1933
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1689 AATTTTGTGTTATGACGAAGATGCGCTCTCCACGACAGGTCAATTTTAAAGTTATCTTTC 1748
QY 1934 cctgcaggataaagtggaaaactctgataaaggttaaaactcaaatcagttattatgtaaccg 1993
Db 1749 AGGTAAATGGAATACTATACAGTTGGAAGTCAAACTTTAATATGTTTTCAGTGTCT 1808
QY 1994 ttgggatt-----tttttaagtggttttaaatattacaatggaaagcatttgcataacca 2047
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1809 CTAAATTTTGAAGAAATTTTGTAGTCTTTACACCTGGGAAAAAGATTTGTAATTTCACTA 1868
QY 2048 ccaaaaatattgtttaattttatgagtagtaattgttagtgcctacgccccatt---- 2103
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1869 TAATAATGTGCTCTTAATTTATAGTAATGTTGTTAATGTTTACATCCCATTTTAA 1928
QY 2104 --aaagcatcaaaatgaatgacatgtagtggtgtgattgacattttagcgaatcaa 2161
Db 1929 AAACAACCAACACACTATTGTTTACAATGTGTGGAGAGTTATGGCATATATATTGAA 1988
QY 2162 gatacctttaataaatatggtgggttactaaagaagtaaacgactcttctctgttattt 2221
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1989 TCAAAAATATATGAACATAAAG--TACCTGTGAAGTAAAAATCCT--TTCTTGAGCATTT 2044
QY 2222 taacactgtcaggaacactgcgaataataaataattactgataaaaaaataaaaaa 2281
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2045 TCACACTTGTAACTGAAATCAGAAATTAATAATTTTACTATAAAGAAGAAATACGA 2104
QY 2282 aaaa 2286
Db 2105 CTATA 2109
```

```
RESULT 7
BC000687
LOCUS
DEFINITION
Homo sapiens, translocating chain-associating membrane protein,
clone MGC:784 IMAGE:3347823, mRNA, complete cds.
ACCESSION
BC000687
VERSION
BC000687.1 GI:12653796
KEYWORDS
MGC.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2722)
Strausberg,R.
Direct Submission
Submitted (15-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaops-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
nisc_mgc@nigri.nih.gov
Contact:
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 5 Row: 1 Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7657654.
Location/Qualifiers
FEATURES
source
1. .2722
/organism="Homo sapiens"
/db_xref="LocusID:23471"
/db_xref="taxon:9606"
/clone="MGC:784 IMAGE:3347823"
/tissue_type="Kidney, renal cell adenocarcinoma"
/clone_lib="NIH_MGC_14"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
92. .1216
/codon_start=1
/product="translocating chain-associating membrane
protein"
/protein_id="AAH00687.1"
/db_xref="GI:12653797"
/translation="MATRKSTKSPVLSHEFVLQNHADIVSCVMVFLGLMFEITA
KASIIIFVTLQYNVTLPAEQATQESVSLYYGIDKLTAVFYMVAIIHVAIOEYML
DKINRMHFSKTRKSPQSAFYLACVWGTFILISNIYISDPTILWRAIPHNH
MTQMKFFYISQALWHLHAPPELYFQTKKEDIPQLVYIGLYLHLAGAYLLNHL
GLVLLVHVEFELHISRLFYSEKRYQKFSWAVLFLGRLTLTLTSLVTGFGGL
ARAENOKLDEFTGNFNLAVRIAIVLASICVQAPMMKFIQFQLRWRHSAFOAPAV
KKKPTVTGRSSKKGTENGVTGTLTNSVADSPRNKKESS"
BASE COUNT 834 a 484 c 480 g 924 t
ORIGIN
Query Match 22.9%; Score 529.2; DB 9; Length 2722;
Best Local Similarity 61.7%; Pred. No. 3.9e-96;
Matches 1081; Conservative 0; Mismatches 598; Indels 73; Gaps 12;
```



```
Qy 436 ccggtgagtagcgggtgagtagcggcgatggggtcccgaaagaagaccgagaaacccc 495
Dy 65 CCTGCAGCGCGGGCGGCCGCCACCATGCGGATTCGCAAGAAAAGCACCAGAGCCCC 124
Qy 496 ccggtgtagccagcaatcatggtgagaaaccacgcgggatatggtctctcggtgggc 555
Dy 125 CCAGTGTGAGCCAGCAATTCGTCTCGAGAATCAGCGGGACATCGTCTCTGTGTGGCG 184
Qy 556 atgttcttgctgggacttatgttcgagggcgacggcgagagatgctgatgcttcctc 615
Dy 185 ATGCTCTTCCGTCGGGCTCATGTTGAGATAACGCGAAAAGCTTCTATCATTTTGTG 244
Qy 616 accctcagcatggaagtcgttgcctccag---cgaaaggctaccctcggtgccagacc 672
Dy 245 ACTCTTCAGTACATGTACACCTCCAGCAACAGAAAGAAAGCTACTGATTCAGTGTC 304
Qy 673 cttaccattatgggttcaagatctgcccacagtggtctctctacatcgtggtggccatc 732
Dy 305 CTTTATTACTATGCGATCAAGATTTGGCTACTGTTTCTTCTACATGCTAGTGGCGATA 364
Qy 733 atcattcagccacattcagagtagctgctagataagctcagccggagagactcagctc 792
Dy 365 ATTATTTCATGCGTAATTCAGAGATATATGTTGGATAAAATTAACAGCGCAATGCAC 424
Qy 793 accaaggcaacaaacaaatgaatgagccggcgagctgagtggttctctacatagt 852
Dy 425 TCCAAAACAAACACAGCAAGTTTAAAGAATCTGGTCACTGAGTGTGCGCTTCTACCT 484
Qy 853 tctggtatctgggtatgatcattctgacctgagaaactcctgtcagagccccactcta 912
Dy 485 GCCTGTGTTGGGCGACATTCATCTCATCTCTGAAACTACATCTCAGACCCAACTATC 544
Qy 913 ttgtggaagtctcagccccacacatgatgacatttcagatgaaattttctacatctca 972
Dy 545 TTATGGAGGGGTTATCCCATTAACCTGATGACATTTCAAAATGAAGTGTTCATATATCA 604
Qy 973 cagtggctactggttctagtttcccgagctctacttccagaaagtcaggaaacaa 1032
Dy 605 CAGCTGGCTTACTGGCTTCACTGCTTTTCTGAACTCTACTTCCAGAAAACCAAAAAGAA 664
Qy 1033 gataccgggttcaactcatcatatcattgacctccacctcttccacattgagggccotat 1092
Dy 665 GATATCTCTCGTCACTGTCTACATGCTTACATGCTTACCTTCCACATTTGCTGGAGCTTAC 724
Qy 1093 ctcttgacttgaaacacctgggctgctgttctgactgctgacatgctgctgagctc 1152
Dy 725 CTTTGAACCTTGAATCATCTAGGACTTGTCTTCTGGGTGCTACATTTATTTTGAATTT 784
Qy 1153 ctctccagcgtgtgcagcctgcttacttcttgggtagcggtaccagaaaggtgtgt 1212
Dy 785 CTTTTCACATTTCCCGCTGTATTATTTTAGCAATGAAAGTATCAGAAAGGATTTTCT 844
Qy 1213 ttgtggcctatcggttttatatccgggagactcgtgacactgattgtctcagtggttaca 1272
Dy 845 CTGTGGGCAGTCTCTTTTGTGGGAAGACTTCTGACTTTAAATCTTTTCTAGTACTGACT 904
Qy 1273 gtaggctctacttggccgggac---aaatcgaaaggaaatgctctctctggtaatctc 1329
Dy 905 GTTGGTTTGGCCTTGCAAGAGCAGAAAATCAGAAAGCTGGATTTTCAGTACAGTGGAACTTC 964
Qy 1330 aatgtgtggcagctaaatcgctgttctgctcctgagttgctcagtatccacaggtgtacata 1389
Dy 965 AATGTGTAGCTGTAGAAATCGCTCTTCTGGCATCCATTTGCGTTACTCAGGCATTTATG 1024
Qy 1390 acatggaccttgacgacctgtggtctcagagatggttagaagatgcgaaatctctatctc 1449
Dy 1025 ATGTGGAAGTTCATTAATTTTCAGCTTCGAAGGTGGAGGAACATCTCTCTTTTCAGGCA 1084
Qy 1450 tgtgggaggaagacggtccagg-----tcgagaaaggacagaa 1491
Dy 1085 CCAGCTGTGGAAGAAACCAACAGTAACTAAAGCGAGATCTTCTTAAAAAAGGAACAGAA 1144
```

```
Qy 1492 aatgagtgaggag-----aatccaaatagaatagattctcccaacaaagaagaagag 1542
Dy 1145 AATGCTGTGAATGGAACATTAATCTCAAATGTAGCAGACTCTCCCGGAATAAAAAAGAG 1204
Qy 1543 aaagc-tccttagcagttgcaagcgaattgattcttaccctcaaggggaatccactcttc 1601
Dy 1205 AAATCTTCATATGAATTAATAAATAATTAATTAATGTGCCCCAAAGAAATCTGCTTCTCA 1264
Qy 1602 ttatgtgtgctctgctgctagagatcttctctcagaaacggtcgtgctcttttgaa 1661
Dy 1265 CTATATCTTTCAGCATTTAGAGATTTTCTGTTCTTGAATAATACAGTCTGTGCTCTTTGAT 1324
Qy 1662 tattcctaagtctgttctctaagtgtttttaagggttttgcagacgctatgagtggggagtg 1721
Dy 1325 TTTTCTGTTTACGCTTTCATGCAATTTT-----TTTAAAGGCAATTTGAGGGGAGGAT 1379
Qy 1722 ggggttaagactaaaccactcagcct-----ctaaatacagctcagaatagtt 1768
Dy 1380 ATTGCTATGAATGAATAAATAATTTTAGCTTAGACTAAGCTACCTGCCCTTCAAAATAGTT 1439
Qy 1769 aacgacacacactcttatttaggttctta-cctcaacgattttccaaacgcttttgt 1827
Dy 1440 TAGGACCACCACCACCATATTTATTTGTTTATTTTGAACATTTTCTAATGATTTGG 1499
Qy 1828 ggtgtagtact-----gcagaattgtgtacataaataat---agtttccctcttccaaatgt 1880
Dy 1500 AGAGAAAACATATTACAAAATTTCCACATATCAGTGATACAATTTCTGCTGTCCACCAAT 1559
Qy 1881 cttatcgaattaaacagctcgtcagaaagtgttctgttcttctcaaatgttctcctgcag 1940
Dy 1560 TTTTATATATACAGAGTGGCTGTCTTAAGAAAGGCCATATTTTAAAGTATATCTTTTCAG 1619
Qy 1941 gataaagtggaatac-----tgataaagggttaaacctcaaatcagttattatgt 1988
Dy 1620 GGTACATGGAATTAATATAAAGTTGGATGTCAAACTTTTAATATGTTTTCAGTGTCTCT 1679
Qy 1989 aacctgtggagttttttaaagtggttttaaatccaattcgaatggaagcattgttcaaaccc 2048
Dy 1680 AATTTTGTGGAAT-TTTTGTAGACTTTACACCTGGGAAAAAAGATTTGTAAATATCACCG 1738
Qy 2049 caaaaatagtgtttaaatttatgagtagtaattgttagtctacgcccccaattaaagc 2108
Dy 1739 AACAAATTGCTGCTTTATTTTATAGTAGTGTATTATTTATATATATACATCCCATTTTAAA 1798
Qy 2109 atcaaaatatga 2120
Dy 1799 ACAAAAACATAA 1810

RESULT 8
AX329795 LOCUS AX329795 1267 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 304 from Patent WO0194629.
ACCESSION AX329795
VERSION AX329795.1 GI:18102773
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrikan,S., Soppet,D.R. and Weaver,Z.
Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 304 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES Location/Qualifiers
Source 1..1267
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 343 a 278 c 275 g 371 t
ORIGIN
```

Query Match	21.4%	Score 493.8	DB 6	Length 1267
Best Local Similarity	65.7%	Pred. No. 4.3e-89		
Matches 752	Conservative 0	Mismatches 387	Indels 6	Gaps
QY	350	gtcgcgcagcgtcgcacccctggcgcacacggcggggaatgggagcccgcgctggg	409	
Db	9	GGCTCGACGGGGCGGTACACGACGACCGGAGGCGGCGCGAGTCGGTGAGCAGCT	68	
QY	410	agcgcgcgagtgatcagcgggtggcgcggtgagtagccggtgagtagcccgcgatgggc	469	
Db	69	GGGAAGAGCAAAACGGGGCGGAGACCTGCAGGCGGCGGCGGCCACCATGGCGA	128	
QY	470	tcgcaagaagaacccaggaaccccccggtgctgtagcaacgaattcatggtgcagaaccc	529	
Db	129	TTGCGAAGAAAGCACAAAGAGACCCCGACGTGTGAGACCCACGAATTCGTCTCGA	188	
QY	530	acgcggatagttctcctcggtggcgatgttcttcgtgctgggaactatggttcgagggca	589	
Db	189	ACGCGGACATCGTCTCCTGTGTGGGATGTCTTCTCTGCTGGGGCTCANTGTTGAG	248	
QY	590	cggccgagatgcagtcggtgttcctccacctgcagcatgagtcggttggtccag---	646	
Db	249	CGGCAAAAGCTTCATCATTTTGTGTACTCTTCAGTACAAATGTCAACCTCCGCA	308	
QY	647	aagggtaccctcgggtccaggaaccttkaccattatgggtcgaagatctggccacag	706	
Db	309	AAGAACAGCTACTGATNCAGTGTCCCTTTATTACTATGGCATCAAGATTGGCTAC	368	
QY	707	tgttcttctacatgctggtggccatcatcttcacgcccaccttcaggagtagctgctag	766	
Db	369	TTTTCTCTACATCTAGTGGCGATAATATTTCATCCGCTAAATTCAGAGTATATGTT	428	
QY	767	ataagctcagccgagactcagcctcacaaaggcaaaacaaatgaatgagggcg	826	
Db	429	ATAAATTACAGCGGAATGCATCTTCCAAAACAAACACAGCAAGTTAATGAATCTG	488	
QY	827	ggcagctgagtggtttctacatagtgctggtatctggggtatgatcatctctggcctcg	886	
Db	489	GTCAGCTTAGTGCCTTCTACCTTTTGGCCTGTGTTGGGGCACATTCATCTCATCTCG	548	
QY	887	agaactccctgtcagaccacactctatttggaaagtcctcagcccaacaatgatgacat	946	
Db	549	AAAACTACATCTCAGACCCAACTATCTTTGAGGAGCTTATCCCCATACCTGATGACAT	608	
QY	947	ttcagatgaaattttctacatctcacagttggcttaactggtttcatagtttcccgagc	1006	
Db	609	TTCAAATGAAGTTTTCATACATATACAGCTGGCTTACGTGGCTTCATGCTTTTCTGAAC	668	
QY	1007	tctacttcagaaagtcaggaaacagatatccccgggtcaactcatctacattggcctcc	1066	
Db	669	TCTACTTCCAGAAACCAAAAAGAGATATTCCTCGTCAGCTGTGCTACATTTGGTCTTT	728	
QY	1067	acctcttccacattggaggggcctactctctgtacttgaaccactggcctgctgtcttc	1126	
Db	729	ACCTTTCACATGTCTGGAGCTTACTTTTGAACCTGAATCATCTAGAGCTGTGTCTTC	788	
QY	1127	tgatgctgactatgctgcagctcctctccagcggtgcagcctgctttactttgggg	1186	
Db	789	TGTTGCTTACATTAATTTTGTGAAATTTCTTTTCCACATTTCCGCGCTGTTTTATTAGCA	848	
QY	1187	atgagcgggtaccagaagggtgtcttgtggcctatcgtgtttatatccgggagactcg	1246	
Db	849	ATGAAAATATCAGAAAGAAATTTCTCTGTGGCAGTCTCTTTTGTGTTTGGGAAGACTTC	908	
QY	1247	tgacactgattgtctcagtggttacagtagggcttcaacttggccgggac---aaatcggg	1303	
Db	909	TGACTTTAATTTCTTTCAGTACTGACTGTGTGGTTTTTGGCTTTCGAAGAGCAGAAAAATCAGA	968	
QY	1304	atgggaatgctctctctgtaatgccaatgttgtggcagctaaaatcgctgtcttctct	1363	
Db	969	AACTGGATTCAGTACTGGAACCTCAATGTGTTAGCTGTAGTAATCGCTGCTTCGCAT	1028	

QY	1364	cgaattgcagttaccagggtgtacataaacatgacaccttgacgaccttgcgttcagagat	1423
Db	1029	CCATTTGGGTTACTCAGGCATTATATGCTGTGAAGTTCATTAAATTTTCAGCTTCGAAGT	1088
QY	1424	ggttagaagatcgcaactctcatcgtctgtgggagaagacagcggtccagggtcgaaaaag	1483
Db	1089	GAGGGACATCTCGCTTTTCAGGCACCAGCTGTGAAGAAGAACCAACAGTAACATAAG	1148
QY	1484	gcaca	1488
Db	1149	GCACA	1153
RESULT	9		
LOCUS	BD005229	1267 bp	linear PAT 31-JAN-2002
DEFINITION	Method of screening of protein.		
ACCESSION	BD005229		
VERSION	BD005229.1	GI:18633190	
KEYWORDS	JP 03075332-T/4.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1267)		
JOURNAL	Todo, N., Okuyama, H., Imamura, M., Ishikawa, H. and Nemoto, K. Method of screening of protein Patent: JP 03075332-T 4 16-FEB-2001; SUMITOMO PHARMACEUTICALS CO LTD, NAOKI TODO, HAJIME OKUYAMA, OTOAKI IMAMURA, HIROKORI ISHIKAWA, KIYOMITSU NEMOTO		
COMMENT	OS Homo sapiens (human) PN JP 03075332-T/4 PD 16-FEB-2001 PF 17-AUG-2000 JP 2000005488 PR 20-AUG-1999 JP 99P 234764 PI NAOKI TODO, HAJIME OKUYAMA, MOTOAKI IMAMURA, HIROKORI ISHIKAWA, PI KIYOMITSU NEMOTO PC C12P1/02, GOIN33/50, C07K14/47, A61K38/17, C12N5/10, C12P21/02// PC (C12P1/02, C12R1:91)		
FEATURES	Key Key Location/Qualifiers		
source	FT CDS Location/Qualifiers		
	1..1267		
BASE COUNT	343 a 278 c 275 g 371 t		
ORIGIN	/organism="Homo sapiens" /db_xref="taxon:9606"		

Query Match	21.4%	Score 493.8	DB 6	Length 1267
Best Local Similarity	65.7%	Pred. No. 4.3e-89		
Matches 752; Conservative	0	Mismatches 387	Indels 6	Gaps 2
QY	350	gtcggcagcggtctgcaccctggcggcagacggcggggatgggagaccgcgcgtggg	409	
Db	9	GGCTGTCAGCGGGCGGTGACACAGCAGCCAGCGGAGCGCGGAGTCGTGAGCAGCT	68	
QY	410	agcgcgcgatgatcagcgggtggcgcgcggtgagtacccgtgagtacccgcggatgggc	469	
Db	69	GGGAAGAGCAGAACCGGGCGGAGCACCTTCAGCGCGGGCGGGCCACCATGCGCA	128	
QY	470	tcgcgaagaagacgcgcaggaaccccccggtgtgtgaccacgaattcatggtcagaacc	529	
Db	129	TTCCGAAGAAAGACCAAGAGCCCCCAGTGTGTAGCCACGAAATTCGTCTCGAGAATC	188	
QY	530	acgcgatatggtctcctcgctgggcatgtcttcgtctgggaattatattcgaggcca	589	
Db	189	ACGGGACATCGTCTCTGTGTGGCAATGCTCTCCCTGCTGGGGCTCATGTTTGAGATAA	248	
QY	590	cggcccgagatgtcgatggttctctcacctcgcagcatggagtcggtgttccccg---	646	

D	b	249	CGGAAAAGCTTCTATCATTTTTTTGTACTTCCTTAGTACAAATGTCACCCTCCACAGCAACAG	308
Q	y	647	aagggtaccctcggggtccaggagacctttacattatggttggaagtcaaatgatcggccacag	706
D	b	309	AAGAACAAGCTACTGAATCAGTGCTCCCTTTATTACTATGGCATCAAGAATTTGGCTACTG	368
Q	y	707	tgttcttcatactgcgtggtggccaatcatcttacccgcaccaatctcaggagtagctgctag	766
D	b	369	TTTTTCTTACATGCTAGTGGCGATAAATTATTCATGCSCGTAATTCAAGAGTATATGTTGG	428
Q	y	767	ataagctcacgccggagactgcagctcacaaaggcaaaaaaaaatgaatatgagggccg	826
D	b	429	ATAAAATTAACAGCGCAATGCACCTCTCCAAACAAAAACAGCAAGATTTAATGAATCTG	488
Q	y	827	ggcagctgagtggtctcatagtgctcgttatctcggggtatgatcatcttgccctctg	886
D	b	489	GTCAGCTTAGTGGGTCTACCTTTTGGCTGTGTGGGGCACATTCATTCATCTCTCTG	548
Q	y	887	agaactgcctgtcagacccccactctatttggaagctctcagccccacaacatgatgacat	946
D	b	549	AAAACACTACATCTCAGACCCAACATCTTATGAGGGCTTATCCCCATTAACCTGATGACAT	608
Q	y	947	ttcagatgaattttctacatctcacagtggtcttacttggttctaatgattcccggagc	1006
D	b	609	TTCAAATGAAGTTTTTCTACATATCACGCTGGCTTACTGGCTTCATGCTTTCTCTGAAC	668
Q	y	1007	tctaectccagaagtcagaaaacagatatccggggtcgaactcatctacattgagctcc	1066
D	b	569	TCFACITTCAGAAAACCAAAGAAAGATATTCTCTCGCAGCTTGCTCFACAAPTGGTCTTT	728
Q	y	1067	acctcttccacattgagggggcctactctctgttaactgaaccacctgggcctgctgttc	1126
D	b	729	ACCTCTCCACATTTGCTGGAGCTTACCTTTTGAACCTTGAATCACTAGSACTTGTCTCTC	788
Q	y	1127	tgatctgcactatgctgtcgagctcctctccagcgtgtgagcctgcttcaacttgggg	1186
D	b	789	TGSTGTCTACATTTATTTGTGTGAATTTCTTTTCCACATTTCCGCCCTGTTTTATTATTAGCA	848
Q	y	1187	atgagcgggtaccagaagggttgtctgtggcctatcgtgtttatatccgggagactcg	1246
D	b	849	ATGAAAAGPATCAGAAGGAATTTCTCTGTGGCAGTTCCTTTTGTGTGGGAGAGATTC	908
Q	y	1247	tgacactgattgtctcagtggtttacagtaggcttcaacttggccgggac---aaatcgga	1303
D	b	909	TGACTTTTAATTTCTTCAGTACTGACTGTGTGGTTTTGGCTTTCRAGACAGAAAATCAGA	968
Q	y	1304	atggaaaatgctctctctgtgtaatgttoaatgtgttggcagctaaaaatcgctgtctctct	1363
D	b	969	AACTGGATTTTTCAGTACTTGGAAACTTCAATGTGTTAGCTGTTAGAAATCGCTGTTCTGC	1028
Q	y	1364	cgaagtgcagtatccagggtgacataaacatggaccttgacacgcttgagcttcagagat	1423
D	b	1029	CCATTTGCGTTTCTCAGGCATTTATGATGTGGAAGTTCATTAATTTTCAGCTCTCGAAGGT	1088
Q	y	1424	ggttagaagatgcgaatcttctatgctgtgaggagaaagacggtccagggtcgsgaaaag	1483
D	b	1089	GGAGGGAACATTTCTGCTTTTCAGGCACCAGCTGTGGAAGNAACCAACACTACTAAG	1148
Q	y	1484	gcaca	1488
D	b	1149	GCAGA	1153

RESULT	10
HSTRAMP	
LOCUS	HSTRAMP
DEFINITION	1267 bp
ACCESSION	linear
VERSION	mRNA
KEYWORDS	1267 bp
SOURCE	Tram gene; Tram protein.
ORGANISM	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	1 (bases 1 to 1267) Gorlich,D., Hartmann,E., Pohn,S. and Rapoport,T.A.
TITLE	A protein of the endoplasmic reticulum involved early in polypeptide translocation
JOURNAL	Nature 357 (6373), 47-52 (1992)
MEDLINE	92244357
REFERENCE	2 (bases 1 to 1267) Hartmann,E.
AUTHORS	Direct Submission
TITLE	Submitted (28-JAN-1992) E. Hartmann, Max-Delbrueck-Centr. f. Molekulare Med., Robert-Roessle-Strasse 10, O-1115 Berlin Buch, FRG
JOURNAL	Location/Qualifiers
FEATURES	1..1267 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_line="Hela" 122..1246 /gene="TRAM" 122..1246 /gene="TRAM" /codon_start=1 /product="TRAM protein" /protein_id="CAA45218.1" /db_xref="GI:37265" /db_xref="SPTREMBL:O15629" /translation="MAIRKSTKSPVLSHSFVQLQNHADIVSCVAMVFLGLMFEITA KASIIFFVLTQNYLTLPATEQATSVSLYYIGIKLATVFFYMLVAIIITHAVIQEYML DKINRMRFYKSKHSFNESGQLSAFVLFACVNGTFLISENIDSPOTILWRAYPHNL MTQNMFFYISQALWHPPELYFOKTKEDIPROLVYIGLYFLHAGAYLNLNLHL GLVLLVHYEFELFHLISRLFYFSENEKYOGFSLMAVLEVLGRLLTLILSLVTGFGCL ARANKQLDSTGNGFNVLRAVLAVLASICVTPQPMWKFINFOLRRWRHSAFQAPAY KKKPTVTKGSSKKNGFNGVRLTISNVADSPRNKKKSS"
BASE COUNT	343 a 278 c 275 g 371 t
ORIGIN	

[illegible]


```
QY 774 cagccggagactcagctcaccacaaagcgaacacacaaattgaatgagccggcgagct 833
Db 447 TRACCGACGGATGCACTTCTCCAAACGAGACACAGCAAGTTTAAACGAGTCTGGTCAGCT 506
QY 834 gagtggtttacatagtgatctgtggtatctctggtggtatgataattctggtcctctgagaactg 893
Db 507 CAGTGCATTCCTACCTTTTGGCTGTGTGTGGGCAGCGTTTCATCTCTATCTCCGAAACCTA 566
QY 894 cctgtcagacacactattgtggaagtctcagcccaacacacatgatgacatttcagat 953
Db 567 TATCTCAGACCCCAACATATCTCTGGAGAGCTATCCCTCGTCAAGCTGTCTACATTTGGCCTTTATCTCTT 746
QY 954 gaaattttctacatctcacagcttggtcttactgtttcatagtttcccgagctctactt 1013
Db 627 GAAGTTTTTCTACATCTCCAGCTGGCGGTACTTGGCTCCAGCATTCCTTGAACCTCTACTT 686
QY 1014 cagaaagtcaaggaacacagatactccgggttcaactcatctacatggcttggcctccactctt 1073
Db 687 CCAGAAAACCAAAAAAGAGATATCCCTCGTCAAGCTGTCTACATTTGGCCTTTATCTCTT 746
QY 1074 ccaattgaggggctctctctgtacttgaaacacacctggcgctgctgtctgtatgct 1133
Db 747 TCAATTTGCTGGAGCCTATCTTGAACCTTGAACCAACCTGGGCTTGTCTCTTCTGTGCT 806
QY 1134 gcactatgctgagctgctctccagcgtgtgcagcgtcttacttggggatgagcg 1193
Db 807 GCACATTTTCGTGAATTTCTTTTCCACATTTCTCGTCATTTTATTTAGTCACGAGAA 866
QY 1194 gtaccgaaggggtgtctgttggcctatcggtttatatccgggagactctgacact 1253
Db 867 GTATCAGAAAGGGTTTCTCTCTGGGCAGTTCCTTTTGTGGGAAGACTTCTGACTCT 926
QY 1254 gattgtcagtggttacagtaggcttccacttggc---cggaacaaatcggaatgaaa 1310
Db 927 AATCTTTTCACTTCTTACCTGTGTGGTTGGTCTCGCAAGACCGAGAAATCAGAAAGCTGGA 986
QY 1311 tgcctctctggaatgtcaatgtgttggcagctaaatcgctgttctgctccctcagttg 1370
Db 987 CTTTACTTGGAAACTTCAACGTGTGTGGCGTTAGGATGCGTGTCTCGCGTCCATTG 1046
QY 1371 cagtatccgggtgtacataacatggaccttgacgacctgtgcttccagagatggttaga 1430
Db 1047 CATCACAAAGCCTTCATGATGTGGAAGTTCATTAACCTTCAGCTTCGGAGGTGGAGGGA 1106
QY 1431 aaatgcgaattcttctgtctgtggggaagacagcaggtccaggtcagaaaagca 1486
Db 1107 ACATTCGCTTCCAGGCCCCACCTGTGAAAGAGAAACCGGCCGTGACCAAGGCA 1162

RESULT 12
BTU19578
LOCUS BTU19578 2387 bp mRNA linear MAM 06-SEP-2000
DEFINITION Bos taurus translation chain-membrane associating protein (tram)
mRNA, partial cds.
ACCESSION U19578
VERSION U19578.1 GI:9971727
KEYWORDS
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 2387)
AUTHORS Clark, T.G., Morris, J., Akamatsu, M., McGraw, R.A. and Ivarie, R.D.
TITLE Cloning and sequence analysis of a bovine tram cDNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2387)
AUTHORS Clark, T.G., Morris, J., Akamatsu, M., McGraw, R.A. and Ivarie, R.D.
TITLE Direct Submission
JOURNAL Submitted (06-JAN-1995) Genetics, University of Georgia, Athens, GA
30602-7223, USA
FEATURES
source
1..2387
```

```

/organism="Bos taurus"
/db_xref="taxon:9913"
<1..2387
/contig="tram"
<1..1077
/codon_start=1
/product="translation chain-membrane associating
protein"
/protein_id="AAG10391.1"
/db_xref="GI:9971727"
/translation="EFVLQNHADIVSCVMVFLGLMEITAKVSIIFVTLQYNVTLPL
ATEQATSEAFLYYGIKDLATVFFVLMVAIIHAIQEVYLDKINRMHFSKTKHSH
FNSGOLSAFYLFSCITWGTILISENVISDPTILWRAYPHNLMTFOKMFYISOLAYW
FHAPPELYFOKTKKEDIPOLVYGLYFLHAGAYLLNLHGLGLVLAHVFVEFLFH
ISRLTFSEDEKYQKGSRLWAVLFLVGLRLLLIILSYLVYSGGLARAENQKLDFTSGNEN
VLAVRTAVLASICITQAFMMKVIENFQLRWRHSAFQAPAVKRPPTVGRSXXKGT
ENGVRITVTSNGADSPNRKESK"
BASE COUNT 675 a 473 c 449 g 788 t 2 others
ORIGIN

Query Match 20.88; Score 481.6; DB 4; Length 2387;
Best Local Similarity 64.0%; Pred. No. 1.4e-86;
Matches 807; Conservative 0; Mismatches 419; Indels 34; Gaps 4;

QY 511 gaattcattggtcagaacacacgagcgtatggtctcctcgtggcgatgttcttctgctg 570
Db 1 GAATTCGTCGTGAGAAATCAGCGGACATCGTCTCTGTGTGGCGATGCTTCTCTGCTG 60
QY 571 ggaattatgttcgagggcagcgagatgctcgatcggttctcctcacccctgcagcatgga 630
Db 61 GGGCTCATGTCGAGATAACAGCAAAAGTTTCTATCATTTTGTACTCTTCAGTACAAT 120
QY 631 gtcgtgtccacg---cggaaggctaccctcggggtccaggaacctttaccattatggg 687
Db 121 GTTACCCTCCCTGCGACAGAGAACAAGCTACTGAATCAGCGTTCCTTTATTACTATGT 180
QY 688 gtcaaatgctggcacagtgcttctacatgctgtggccatcatcatcagcgcacc 747
Db 181 ATCAAGATTGGTCAGATTTTCTTACATGCTAGTGGCGATANTTATTATTCATGCCATA 240
QY 748 attcaggagtacgtgtcagataaagctcagcgagactgcagctcaccaaaaggacacaa 807
Db 241 ATTCAAGATATGTGTTGGATAAATTAACAGCGGAATGCATTTCTCCAAAACGAAACAC 300
QY 808 aacaaattgaaatgagggcagcagtggtgttctacatagtgctcgtggtatctggtggt 867
Db 301 AGCAAGTTTAAATGAATCTGGTTCAGCTTAGTGCATTTCTACCTTGTGTATTGGGCG 360
QY 868 atgatacttctgacctgagaactgctgtcagacccacctctatgtggaagtctcag 927
Db 361 ACATTCATCTCATATCTGAAACTATCATCTCAGACCCGACTATTTGTGGAGGCTTAT 420
QY 928 ccccaacaatgatgacatttcagatgaaattttctacatctcacagttggttactgg 987
Db 421 CCCCATAACTGATGACATTTTCAGATGAAGTTTTTCTACATATACACAGTGGCATCTGG 480
QY 988 ttctaatgtttcccgagctctactctccagaagctcaggaacacagatatccccgggtcaa 1047
Db 481 TTTTCATGCTTTTCTCGAATCTACTTCCAGAAAACCAAAAAGAGATATTTCTCTCGTCAG 540
QY 1048 ctactctaatgttgacctccacctcttccacattggaggggacctatctcttgaactgaac 1107
Db 541 CTCGCTCATACCGGCTTTACCTCTTTCACATTCGAGCTTACCTTTTAAACTTGAAT 600
QY 1108 cactggcgctgctgtctgtgactgcactatgctgcagctccctctcctccagcgtgctg 1167
Db 601 CATCTAGGACTGTGGCTTTTGGTGTCTACATATTTTGTGCGAGTTCCTTTTCCACATTTCC 660
QY 1168 agcctgtttactttggggatgagcgggtaccagaaggggtgttcttctgtgacctatgctg 1227
Db 661 CGCCTATTTTACTTTACTGACGAAAAGTATCAGAAAGGATTTTCTCTCTGGGCGACTTCTG 720
```



```
Db 969 ATGTGGAGTTTCATTAACCTTCCAGCTTGGAGGTTGGAGGGAACATCTTGCTTCCAGGCC 1028
Qy 1450 tgtgggaggaagacagcgttcaggtcgagaaagcga 1486
Db 1029 CCACCTGTGAAGGAACCGCGCTGACCAAGGCA 1065

RESULT 14
AY029534 1338 bp mRNA linear VRT 02-MAY-2001
LOCUS Xenopus laevis Traml mRNA, complete cds.
DEFINITION AY029534
ACCESSION AY029534.1 GI:13936286
VERSION 1
KEYWORDS African clawed frog.
SOURCE Xenopus laevis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 1338)
Hartmann, E.
Direct Submission
Submitted (10-APR-2001) Biologie, University Lubeck, Ratzeburger
Allee 160, Lubeck 23538, Germany
Location/Qualifiers
1..1338
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="RZPD A0180202"
16..1137
/codon_start=1
/product="Traml"
/protein_id="AAK40302.1"
/db_xref="GI:13936287"
/translation="MGIRKNSKTPPVLSHEFTIQNHADIVSLCMVFLGLMFETIA
KVAVFVTLQYNTIPVEGLGPTSLYHYGIKDMATVFFYMLVAIILHVAIOEYVLD
KINRMHFSKTHSKNESQLSAFYFSCIGWASIIVSSENFSDPSIMKGYPHTVF
PFOMKFYISOLAYFHARPELFYFORKKEDIIPOLVYIGLYLFIHGIAGVYLNRLG
LVLLVLYHVEFVLFHMSRFLYFSNRYKQFTWAVLFLGRLLILLVLTGVFGFLA
RAENQELDSNGFNFIILIRITVLIASICITQAFMMKFINFQLRMRHSSPQFSOR
KKATSAKGASRKREKGVNGVTSTNSGASPRSRKEKHS"
BASE COUNT 362 a 321 c 267 g 388 t
ORIGIN

Query Match 19.08; Score 438.6; DB 5; Length 1338;
Best local Similarity 63.28; Pred. No. 5.6e-78;
Matches 691; Conservative 0; Mismatches 399; Indels 3; Gaps 1;

Qy 462 catgggctcgcaagaagcagcaggaaccccggtgctgagccacgaattcatggt 521
Db 15 CATGGGCATCGTAAGAAGAACACAGACACCCCGCGTGTACGCCAGGATTCATTAT 74

Qy 522 gcaagaaccagcgataggtctctcggtggcgatgcttctgctggtgggaactatgtt 581
Db 75 CCAGAACCATGCGGACATGCTCTCTGCTGGCCATGCTCTCTGCTCGGCCCTATGTT 134

Qy 582 cgaaggcacggccagatgctgctgcttccaccctcgagcagatgagtcggtgtccc 641
Db 135 CGAGATTACAGCAAGGTGGCTGTCATGTTGTGCACATCCAGTACAAATGTCAACATTC 194

Qy 642 agcggaaggctaccctcggtcgagaccctttaccattatgggtcgaagatcggc 701
Db 195 AGTAGAAGGGTGCTTGGAGAACCCACTCGCTACCATTTATGGCATCAAGATATGGC 254

Qy 702 cacagtgtctctacatgctggtggccatcatcattcaecgccaccattcaggatcgt 761
Db 255 CACTGTATTTTACATGCTGGGGCAATAATACTACTAGTCAGTTCATCCAGAAATACGT 314

Qy 762 gctagataagctcagcggagactgcagctcaccacaaagggcaacaaacaaatgaatga 821
Db 315 ATTAGATAAAATCAACCGCGGTATGCATTTTTCACAGCAAAACACAGCAAGTTTAATGA 374
```

```
Qy 822 ggccgggagcagtgagtggtttctacatagtgctcgtgatatctgggtgatgatcattctgac 881
Db 375 GTCGACAGCTAAGTCGATTTTACCTGTTCTCTTGGCATTTTGGGGAGCAAGCATATTCTT 434
Qy 882 ctctgagaactgcctgctcagacccaccctctattgtggaaagtctcagcccccacaaatgat 941
Db 435 CTCAGAGAATTATTTTTCGGATCCAATAAGCCTGTGGAAAGGGTATCCTTCACACATACATT 494
Qy 942 gacattcagatgaattttttctacatctcacagttggcttactggttttctagtttctcc 1001
Db 495 TCCATTCAGATGAAGTTCTTCTACATTTCCAGCTGGCTTATTTGGTTTCATGCTTTTCC 554
Qy 1002 ggagctctacttccagaagtcaggaacaaagatatcccggttcaactcatcatcattgg 1061
Db 555 AGAGCTCTACTTTCAGAAAAACGAAAGAGACATTCACAGGAGCTGGTCTACATTGG 614
Qy 1062 cctcacctcttccacatggaggggacctatctcttctgttacttgaaccaccctggcctgct 1121
Db 615 TTTATACCTCTTCCATATTATAGAGACTTATGTTCTCAACCTCAATCGCCTGGCTTGGT 674
Qy 1122 gcttcgagtcgactatgctcagctcctccagcgttccagcgtgagcgtcgttactt 1181
Db 675 ACTGCTAGTTCTGCAATTACTTTGTAGAAATTCCTGTTCATATGTCACGCTCTGTTTACTT 734
Qy 1182 tgggatgagcgttaccagaagggttcttctgtggcctatcgtgttttatatccgggag 1241
Db 735 CAGCAATGAGATACACAGAGGGTTTACTGTGTGGCGCTTCTCTTTGTCTCTTGGGG 794
Qy 1242 actcgtgacactgattgtcagtggtttacagtaggggttccacttggcgggac---aaa 1298
Db 795 CTTCTCTCACTCTTATCTCTTCTGTTCTTACTCTCGCTTTGGACTTGCACGGCTGAAAA 854
Qy 1299 tcggaatggaaatgctctctctgtaagtcaatgtgttggcagctaaatcgctgttct 1358
Db 855 TCAAGAGTTGGATTCGAGTAATGGAAATTTCAACATTTTGGCCATTCAGAAATCACCGTGT 914
Qy 1359 gtccctcaggttcagatccaggtgtacataacatggaccttgagaccgtctggcttca 1418
Db 915 GGCATCTATTTGCATCACCAGGCATTTATGATGTGGAAATTCATTAATTTCCAGCTGAG 974
Qy 1419 gagatggttagaagatcgcaatcttcctgctgtgtgtgggaggaagacggtccaggtc 1478
Db 975 ACGATGGAGAGAGCATTCATCTCCTCAACCCCTCCTTCAAGGGAAGAGGCGACCTCTGC 1034
Qy 1479 aaaggccacagaaaaatggagtgagagaatcccaatagaatagattctccaccacaaagaa 1538
Db 1035 TAAAGGCAAGCTTCACCTAAAGAAAAAGAAATGGAGTGAATGGAACAGTAAACATCAAA 1094
Qy 1539 agagaaagctcct 1551
Db 1095 CGAGCAGATTTCT 1107

RESULT 15
AY029529 1453 bp mRNA linear VRT 02-MAY-2001
LOCUS Danio rerio Traml mRNA, complete cds.
DEFINITION Danio rerio
ACCESSION AY029529
VERSION AY029529.1 GI:13936276
KEYWORDS zebrafish.
SOURCE Danio rerio
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 1453)
Hartmann, E.
Direct Submission
Submitted (09-APR-2001) Biologie, University Lubeck, Ratzeburger
Allee 160, Lubeck 23538, Germany
Location/Qualifiers
1..1453
/organism="Danio rerio"
```

```
/db_xref="taxon:7955"
/clone="MPMGp609N2214Q8"
79..1188
/codon_start=1
/product="Tram1"
/protein_id="AAK40297.1"
/db_xref="GI:13936277"
/translation="MGIRKTKNPPVLSHFVIONHADVSCVAMVFLGLMFEITS
KVAVLFIQVYNTIPANGPEEATVNFHYGLKDVATIFFYMLVAIIHAIIOEVL
DKINRKHFSKTHKSNESGOLSAPYLEFSLWASILGSENILSNPVLWDGYPHFL
MQFQMKFYICQGLGFWLHAPELIFQKAKKEDIPLCVIISLILVHIAGVAINLRLL
GLVLLVHYFLIEFHVSLRILYFSNEERQSGTVMWVLFVLGRLLTSLSVLTGVGL
AGAERQGLNLAEGSNVLFVRVTLAAICITQAFMWKFIINFQLRRWREQAQTQLKK
KSSSKSKSKKANGVSGVSGADSPRARKEKS"
BASE COUNT      371 a   340 c   323 g   419 t
ORIGIN

Query Match      16.9%; Score 390.8; DB 5; Length 1453;
Best Local Similarity 62.7%; Pred. No. 2.3e-68;
Matches 643; Conservative 0; Mismatches 377; Indels 6; Gaps 2;

QY 463 atggggtctgcgaagaagacgcaggaacccccccgggtgctgagccacgaattcatggtg 522
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 79 ATGGGCATCCGAAAAAGACCACCAAGAACCCCGCGGTGCTCAGCCATGAGTTTGTGATC 138
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 523 cagaacacgcggtatgctctcgtcgtgggcatgtcttctgctggtgacttatgttc 582
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 139 CAGAACACGCGGATATTTGTCTCTGTCGTATGGTGTCTCTGCGTCTTAATGTTTT 198
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 583 gagggcagcgcgagatgctgatctgttctctcaccctgcagcagtgagtcgtgtccca 642
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 199 GAGATTACATCAAAAGGTAGCGGTTTTGTTCAATACTGTGCAATACAAACGTCACGATTCCA 258
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 643 gcggaaaggctaccctcgggtccaggaacccctttac---cattatgggtcacaagatctg 699
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 259 GCAAAATGGAGGTCACAGGAAGACGACAGTGAATTACTTTCAATTATGCTCTGAAAGATGTG 318
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 700 gccacagtgtctctacatgctgggtggccatcatcattcacgccaccattcaggagtac 759
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 319 GCCACCATTCTTTTACATGTTGGTGGCCATTTATCATGCACGCCCATTTATTCAGGAATAT 378
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 760 gtgctagataagctcagcgggagactgcagctcaccaaaggcaaaacaaaataatgaat 819
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 379 GTGCTGGATAAAATCAACAGGAAGATGCACTTCTCTAAAAACCAAGCACAGCAAGTTCAAT 438
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 820 gaggcgggcagctgagtgtgttctacatagtgctcgtggtatctcgtgggtatgatcattctg 879
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 439 GAGTCGGGTGAGTGAGTGCCTTCTACCTCTTCTCTGCTGTGGGGAGCCAGCATTCCTT 498
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 880 gcccttgagaactgcctgcagaccctcactctattgtggaagtctcagccccaacaacatg 939
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 499 GGTCTTGAGAAACATCTCTCAAAACCTGTTCAGCCTGTGGGATGGTTACCCACACACCCTA 558
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 940 atgacatttcagatgaataattttctacatctcacagtttggttactcgtggtttcattagtttc 999
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 559 ATGCAGTTCCAGATGAAGTTCTACTACATCTGTCACTGCTCGGCTACTGGCTTCATGCAATC 618
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1000 ccgagactctacttcagaaaagtcagaaacaagatatccccgggtcacaactcattcatt 1059
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 619 CCAGAGCTTTACTTCCAGAAAGCCCAAGAAAGAGACATTTCTCGCCAGCTTGTTTATATA 678
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1060 ggcctccacctcttcacattgaggggacctatcttcttacttgaaccacctgggacctg 1119
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 679 AGCCTTTACTTGGTCCACATCGCTGGAGCCTACGGCTCTGAATCTGAACCGGTTGGGTTTG 738
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1120 ctgcttctgctgactgactgctgctgcagctcctctccagcggtgtgcagcctgctttac 1179
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 739 GTGTTGCTGGTCTGCAATTACTTTGTTGAGCTCATCTTCCATGTGTCAGACTCATTTTAC 798
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1180 ttggggagtagcgggtaccagaaaagggttcttcttggcctcattcgttttatatccggg 1239
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 799 TTCAGCAATGAGGAAAGGAGAGATGGGTTTACAGTGTGGGCTGTTCTTGTGTTGCTGGGC 858
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 1240 agactcgtgacactgattgtctcagtggttacagtagggcttcacttggccgggaca--- 1296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 859 CGGCTGTTGACTCTGTCTCTGTCTTCTTACTGTGGGCTTCGGCCTTGACAGGACGAGAG 918
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1297 aatcggaatggaaatgctctctctcgttgtaatgtcaatgtgttgagcagctaaaaatcgctgtt 1356
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 919 AAACAGGGTTTGAATCTTGCAGAGGGAAGCTTTAATGCTCCTGTTGTGCGAGTGACCGTT 978
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1357 ctgtcctcagattgcagtatcccagggtgtacataacatggaccttgacgaccgctctgctt 1416
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 979 CTGGCTGCCATATGCCATCACTCAAGCTTTTATGATGTGGAATTTATCAACTTCCAGTTG 1038
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1417 cagagatggttagaagatgcgaatcttctcattgtgtggtggaggaagacggtccaggctcg 1476
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1039 CGACGATGGAGAGACAAAGCACAGACTCAGACCTTAAAGAGAATAATCATCTCTCTTAAG 1098
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1477 agaaaa 1482
|||||
DB 1099 AGCAAA 1104
```

Search completed: September 6, 2002, 14:46:34
Job time: 36398 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 12:13:06 ; Search time 495.6 Seconds
(without alignments)
8006.032 Million cell updates/sec

Title: US-09-807-470-1
Perfect score: 2311
Sequence: 1 agagagagagagagagag.....aaaaaaaaaaaaaaaaaaaaa 2311

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneq-emb/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2311	100.0	2311	21 AAA38012	Rat WAR-1 nucleoti
2	1092	47.3	1092	22 AAF74781	Rat WAR-1 encoding
3	802.2	34.7	2288	21 AAA38013	Human WAR-1 nucleo
4	770.6	33.3	1835	23 AAS76421	DNA encoding novel
5	718.2	31.1	1736	22 AAK94181	Human full-length
6	695.2	30.1	1288	22 AAF74782	Human WAR-1 encodi
7	647.8	28.0	1110	22 AAF74780	Human WAR-1 encodi
8	532	23.0	2833	21 AAC77810	Human cancer assoc
9	517.8	22.4	2756	24 AAS62269	CDNA sequence #56

10	477.6	20.7	1517	24 AAS18576	cDNA encoding huma
11	414	17.9	1120	20 AA233531	Human prostate can
12	372.2	16.1	802	22 AAK91674	Human CDNA 5'-end
13	372.2	16.1	802	22 AAK93837	Human CDNA clone r
c 14	345	14.9	642	22 AAS47389	Human breast cance
c 15	345	14.9	642	22 AAF17959	Human breast cance
16	260.6	11.3	1380	22 AAL01678	Human reproductive
17	212.6	9.2	1526	23 ABL18595	Drosophila melanog
18	212.6	9.2	1549	23 ABL18061	Drosophila melanog
19	212.6	9.2	2153	23 ABL21096	Drosophila melanog
20	212.6	9.2	3346	23 ABL21098	Drosophila melanog
21	212.6	9.2	3618	23 ABL18060	Drosophila melanog
22	212.6	9.2	4146	23 ABL18594	Drosophila melanog
23	212.6	9.2	4775	23 ABL21091	Drosophila melanog
24	212.6	9.2	5055	23 ABL21094	Drosophila melanog
25	212.6	9.2	11287	23 ABL21090	Drosophila melanog
26	202	8.7	472	21 AAC00195	Human secreted pro
c 27	109.4	4.7	545	22 AAK92492	Human CDNA 3'-end
28	78	3.4	271	22 AAL23061	Human breast cance
c 29	77.2	3.3	422	22 AAL15924	Human breast cance
30	76.6	3.3	272	22 AAL14199	Human breast cance
31	69.2	3.0	402	22 AA184805	Membrane-bound pro
32	68.2	3.0	3265	21 AA265095	Human DNA encoding
33	68.2	3.0	3265	22 AAS46053	Human DNA encoding
34	68.2	3.0	3265	22 AAF92092	Human PRO1124 cDNA
35	68.2	3.0	3265	22 AAF44241	Human PRO1124 (UNQ
36	68	2.9	1119	21 AAC60033	Human secreted pro
c 37	67.8	2.9	18011	24 ABL32034	Human immune syste
38	67.6	2.9	1254	22 AAS25886	Human CDNA encodin
c 39	67.4	2.9	852	22 AAL19845	Human breast cance
c 40	67	2.9	545	22 AAH70126	Human cervical can
41	67	2.9	958	22 AAC92233	Apoeaquorin-encodi
42	67	2.9	958	24 AAD22188	Aequorea victoria
43	67	2.9	959	9 AAN81534	PAQ440 aequorin ge
c 44	66.8	2.9	310	22 AAH71505	Human cervical can
45	66.6	2.9	1952	21 AAC59577	Human secreted pro

ALIGNMENTS

RESULT 1
AAA38012
ID AAA38012 standard; DNA; 2311 BP.
XX
AC AAA38012;
XX
XX 22-AUG-2000 (first entry)
DT
DE Rat WAR-1 nucleotide sequence.
XX
XX Endoplasmic reticulum; WAR-1; cancer cell proliferation inhibitor;
KW diagnosis; cancer; sarcoma; rat; ds.
XX
OS Rattus norvegicus.
XX
XX WO200022123-A1.
PN
XX 20-APR-2000.
PD
XX 13-OCT-1999; 99WO-JP05631.
PF
XX 13-OCT-1998; 98JP-0290711.
PR
XX (SUMU) SUMITOMO PHARM CO LTD.
PA Tohdoh N, Yoshima T, Komiya K, Tojo S, Nemoto K, Ishikawa H;
PI Okuyama H;
XX WPI; 2000-317980/27.
DR P-PSDB; AAY98146.
XX
PT Endoplasmic reticulum protein WAR-1 which inhibits cancer cell

Db 1861 agtttctgcttccaagtgttcttcttattcgaattaaacaagctgctagcaaaagtggtttgtt 1920
Qy 1921 ttctcaatgttctctcagcagataaagtggaaatctgtataaaaggttaaaactcaaatcag 1980
Db 1921 ttctcaatgttctctcagcagataaagtggaaatctgtataaaaggttaaaactcaaatcag 1980
Qy 1981 tattatgaaccgttgaggatttttttaaaagtgttttaaaattacaatggaagcatttgt 2040
Db 1981 tattatgaaccgttgaggatttttttaaaagtgttttaaaattacaatggaagcatttgt 2040
Qy 2041 caaacaccacaaatatgtgttttaattttatagtagtaattgttagtcttaagcctccc 2100
Db 2041 caaacaccacaaatatgtgttttaattttatagtagtaattgttagtcttaagcctccc 2100
Qy 2101 attaaacatcaaaatatgaatgacatgtgtgtgtgatatgtgacatttagcgaatca 2160
Db 2101 attaaacatcaaaatatgaatgacatgtgtgtgtgatatgtgacatttagcgaatca 2160
Qy 2161 agataccttttaataaataatggtggttactaaagaagttaaacgactttctctgtttatt 2220
Db 2161 agataccttttaataaataatggtggttactaaagaagttaaacgactttctctgtttatt 2220
Qy 2221 ttaaacactgtacagaaaactcgcgaataatataattactgaaaaaataaaaaa 2280
Db 2221 ttaaacactgtacagaaaactcgcgaataatataattactgaaaaaataaaaaa 2280
Qy 2281 aaaaaaaaaaaaaaaaaaaaaa 2311
Db 2281 aaaaaaaaaaaaaaaaaaaaaa 2311

RESULT 2

AAE74781
ID AAE74781 standard; cDNA; 1092 BP.
XX
AC AAE74781;
XX
DT 17-MAY-2001 (first entry)
XX
DE Rat WAR-1 encoding cDNA sequence SEQ ID NO:4.
XX
KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
KW endoplasmic reticulum membrane transportation; secretory protein;
KW cell membrane protein; cytosolic; CNS active; antiallergic; cancer;
KW antirheumatic; nervous system disorder; immune disorder; allergy;
KW rheumatism; skeletal disorder; ss.
XX
OS Rattus sp.
XX
FH KEY Location/Qualifiers
FT CDS 1..1092
FT /*Cag= a
FT /product= "WAR-1"
XX
XX WO200114582-A1.
XX
XX 01-MAR-2001.
XX
XX 17-AUG-2000; 2000WO-JP05488.
XX
XX 20-AUG-1999; 99JP-0234764.
XX
XX (SUMO) SUMITOMO PHARM CO LTD.
XX
XX Tohdoh N, Okuyama H, Imamura M, Ishikawa H, Nemoto K;
XX WPI: 2001-202940/20.
XX P-PSDB; AAE70696.
XX
XX Transformation of a cell with separate vectors expressing the sense and
XX antisense strands of WAR-1 DNA for screening secretory and membrane
XX proteins expressed by the cell -

PS Claim 3; Page 65-68; 79pp; Japanese.
XX
CC The present invention describes a screening method for secretory and
CC membrane proteins consisting of transformation of a cell with separate
CC expression vectors for the sense and antisense RNA of DNA encoding an
CC endoplasmic reticulum membrane protein participating in endoplasmic
CC reticulum transport of proteins. Also described are: (1) secretory and
CC cell membrane proteins identified by the screening method; (2) drug
CC compositions containing these proteins; (3) host cells transformed by
CC the separate expression vectors of the method; and (4) the preparation
CC of secretory and cell membrane proteins by culture of the transformants.
CC The method can be used for the identification and preparation of
CC proteins for use in the treatment and prevention of diseases such as
CC cancer, disorders of the nervous system, immune disorders (including
CC allergies and rheumatism) and skeletal disorders. The present sequence
CC encodes a specifically claimed rat WAR-1 protein from the present
XX invention.
SQ Sequence 1092 BP; 260 A; 268 C; 288 G; 276 T; 0 other;
Query Match 47.3%; Score 1092; DB 22; Length 1092;
Best Local Similarity 100.0%; Pred. No. 2.4e-202;
Matches 1092; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 463 atggggctccgcaagaacacccaggaaccccccgggtgtcagccacgaattcattcggtg 522
Db 1 atggggctccgcaagaacacccaggaaccccccgggtgtcagccacgaattcattcggtg 60
Qy 523 cagaacccacgggatgtgtctctcggtggtcgtggttctctgtggtggtggttcttc 582
Db 61 cagaacccacgggatgtgtctctcggtggtcgtggttctctgtggtggttcttc 120
Qy 583 gagggcacggccgagatgtcgtatgttctctcaccctcagcagcatggagtcgttcccca 642
Db 121 gagggcacggccgagatgtcgtatgttctctcaccctcagcagcatggagtcgttcccca 180
Qy 643 gcgggaaggtaccctcggtggtccagaccctttaccattatgggtcctcaaatcctggcc 702
Db 181 gcgggaaggtaccctcggtggtccagaccctttaccattatgggtcctcaaatcctggcc 240
Qy 703 acagtgttcttctacatgctgtggtggtccatcatcattcagccaccattcagagtagctg 762
Db 241 acagtgttcttctacatgctgtggtggtccatcatcattcagccaccattcagagtagctg 300
Qy 763 ctagataagctcagccggagactgcagctcaccgaaggcaacaaacaaatgaatgag 822
Db 301 ctagataagctcagccggagactgcagctcaccgaaggcaacaaacaaatgaatgag 360
Qy 823 gccgggcagctgagtggttctacatagtgctgtgtatctggtgtatggtatgattcctggcc 882
Db 361 gccgggcagctgagtggttctacatagtgctgtgtatctggtgtatggtatgattcctggcc 420
Qy 883 tctgagaactgctgtcagaccctctctgtggaagtctcagcccccacacacatgatg 942
Db 421 tctgagaactgctgtcagaccctctctgtggaagtctcagcccccacacacatgatg 480
Qy 943 acatttcagatgaaatttttctacatctcagagtggtggttctggtttcattagttcccg 1002
Db 481 acatttcagatgaaatttttctacatctcagagtggtggttctggtttcattagttcccg 540
Qy 1003 gagctctacttccagaagtcaggaaaagataccccgggtcacaactcattcattggc 1062
Db 541 gagctctacttccagaagtcaggaaaagataccccgggtcacaactcattcattggc 600
Qy 1063 ctccacctctccacatggaggggcctctctgttacttgaaccacctgggctgctg 1122
Db 601 ctccacctctccacatggaggggcctctctgttacttgaaccacctgggctgctg 660
Qy 1123 ctcttgatgctgacatgtctgcagctctctccagctgtgcagctggttacttt 1182
Db 661 ctcttgatgctgacatgtctgcagctctctccagctgtgcagctggttacttt 720

[illegible]

```
QY 523 cagaacacgcgatatggtctcctcgtggtggcgatgttcttcgtctgctgggacctatgttc 582
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 cagaatcatcgacatcgctcctcgtcgtgggatgttcttcgtcgtggcgctgtgttc 120

QY 583 gagggcacgcccagatgtcgatcggttctcctcacctcgacgatggagtcgtgtcc-- 640
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 gagggacagcagaagcatccatcggttctcactccttcagcacagtggtcgtccct 180

QY 641 -cagcggaaagggtacacctcggggtccagacaccttaccattatgggtcacaagatctg 699
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 gcagcagagaaacagcccggtctcaagctccctctattattatgggtccaagatttg 240

QY 700 gccacagtggttcttcacatgctgggtggccatcatcatcagccaccatcaggagtac 759
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 gccacggttcttcacatgctgggtggccaatcatattcatcgcacaaatccaggaatat 300

QY 760 gtctagataagctcagccggagactgcagctcaccaaaaggcacaacaaacaaatgaat 819
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 gtgttgataaaaattacaagaaatgcagttccacaaagcgaacaaacaaagtttaac 360

QY 820 gaggcggcgagctgagtggttctacatagtgctcgttatctgggtatgatactctg 879
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 gagtctggctcagtggttctcactcttttctctgtatttggggcaccatcatttta 420

QY 880 gcctctgagaactgcctgtcagaccccaactctatttgggaagctcagcccccacaacatg 939
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 atctctgaaaactgcctgtcagacccaactcttatatggaaggctcgtccccatagcatg 480

QY 940 atgacatttcagatgaattttctacatcctacatcagctcagcttggttctatagtttc 999
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 atgacatttcagatgaattttctacatcctcagcttggttctatggttctatgctttt 540

QY 1000 ccggagctctactccagaaagtcagaaacaaagatatcccggtcacaactcatctacatt 1059
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 cctgaactctactccagaaacaaacaaacaaagacatccctcgtcacaactgtctacatt 600

QY 1060 ggctccacactctccacattgagggggacctatctcttgaacttgaacacacctggccctg 1119
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 ggtctccactctccacattactgagcttatctctgttaacttccacatgctggcctgtttac 720

QY 1120 ctgctctcagctgcactatgctcagagctcctcctcagctgtgcagcctgctttac 1179
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 ctctcttggtcagcttatctgttgaattacttccacatgctggcctgtgtttac 720

QY 1180 ttggggatgagcgttaccaagaagggtgtgttcttggtgacctcgtgtttatatccggg 1239
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 tttagtgataaagtaccagaaggcatatctctgtgggccaattgtgttatctctgggt 780

QY 1240 agactcgtgacactgattgtctcagtggttacagtgagggtcctcacttggccgg---gaca 1296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 agacttggactttaattgtttccgtactcaactgttgggtttcacctggctggatcgacg 840

QY 1297 aatcggaatggaatcctctcgtgttaattgtaaatgtgttgccagctaaaatcgctgtt 1356
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 aatcggaactcgtgatcccttactgctgaaatgtaaatgtgttgccagctaaaatgtgtt 900

QY 1357 ctgtctcagtggtcagtatccagggtgacataaacatgaccttgacgacccgtctgctt 1416
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 ctgtctcagtggtcagtatccaggctcgttaaacatggaaacttaattactctctgctt 960

QY 1417 cagagatggttagaagatcgaaatcttcatgtctgttggaggaa-----gaga 1464
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 cagagtggttagaagattcttaattcaggcctcatgtatgaaaagaacggtcgaga 1020

QY 1465 cgggtccaggtcgagaaaaagccacagaaaaatgagtgagagaatccaaatagatatctt 1524
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 tcttctaaaaaagacagaaacgagtggtggagtggaacttcaaatagatgagactgt 1080

QY 1525 ccaccaagaagaagagaaagctcttagcaggttgcaagcgaattgattcttaccctcca 1584
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 ccgccaagaaggagaaagaaatcttcaataatcttgcagcgcattgatttaagtctgca 1140
```

```
QY 1585 agggaatccactctcttattatgtgtctcgtgctagaga-ttttctgtttcttcagaa 1643
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 aaggaat---ctgctcttggaggttctcttctcgtcactagagatttgtgttttgaaaa 1197

QY 1644 cgggtcgtgcttttgaatattgctaattgtattgtcttaattgtgtttt----- 1691
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1198 tagttcgtcgtctcgtttttgttattgaactgttctatgtttttaaagacatttg 1257

QY 1692 -----aagggttttcagacgtatgagtggggagtggttaagactaaacaccacag 1744
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1258 aggggagagagattattatgaatgggaaaaaagattttgttgagactaaatactcat 1317

QY 1745 ccttaaatcacgtcagaatagtttaacgg-----accacatctatttagttaggttctt 1800
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1318 cgtcaaaaataatgcacaaatagtttggggatcaccactatatttgttttgattttaa 1377

QY 1801 acctcaacgatttttccaaaagcttttgggtgagactgcagaaattgtgtacataaataat 1860
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1378 ctttcaacattttcctaattgatttcagagataactgcacaattttgcatacaatgat 1437

QY 1861 agt-----ttctcgtcttccaatgttctttatcgaattaaacagctctctagcaagtggt 1915
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1438 actggttcttactcccacaggtgtttcataatactaaacaagatggtctctcctagcaaga 1497

QY 1916 ttgttttctcaatgttctcctgcaggataaaagtggaaaactctgataaaggtttaactcaa 1975
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1498 ttatgtgtttaatgctgttctgttgggttaaaa--taaaagtacgaaaaaggtggaagtcaa 1555

QY 1976 atcagttattatgaacggttgggatttttttaaaagtgttttaaattttaacaatggaaaaga 2035
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1556 atcagttattgtaattgttgaataattttttaagaacttaacaactcagaaaaagattg 1615

QY 2036 ttgtcaaacacacca-aaaatatgttttaattttatgagtagtaattgttagtgcttac 2094
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1616 ctgagactcaccacaaataaataatgttctttattttacaggtagtgattatagtgctca 1675

QY 2095 gccoccataaagcatcaaaaataatgatatgacatgtgtggtgatatgtgacatttagc 2154
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1676 tccccatttaaaaaaacacagtaactaattgggtaacacatatggag-----gtttgc 1726

QY 2155 gaatcaagatacctttaataaatatggtgggttaactaaagaagtaaacgacttctctcgt 2214
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1727 tgcataataattgcatcaaaaataatcatttaataataaaattataaatcattctcgt 1786

QY 2215 ttattttaaacactgttacaggaaaaactcgcaaaatttaataattactgaaaa 2267
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1787 tccatt---ccactgtaaatgggaattcattgaattaaataattactttaaa 1835

RESULT 5
AAK94181
ID AAK94181 standard; cDNA; 1736 BP.
XX
AC AAK94181;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 2724.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
```


(HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX WPI; 2001-524255/58.
 DR P-PSDB; AAM93265.

XX 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX

PS Claim 8; SEQ ID NO 2724; 1380pp + sequence listing; English.

XX The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesising the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a full length
 CC human cDNA of the invention.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 1736 BP; 448 A; 358 C; 390 G; 540 T; 0 other;

Query Match 31.1%; Score 718.2; DB 22; Length 1736;
 Best Local Similarity 72.2%; Pred. No. 6.9e-130;
 Matches 1062; Conservative 0; Mismatches 363; Indels 46; Gaps 8;

QY 435 gccggtgagtagcggtagtcacgcggcgtggtggtcgcgaagaagacccaggaacc 494
 DB 157 gccggtgagtagtcggga-agcgcacccatggggtcctcgaagaagagcaccagaaacc 215
 QY 495 ccgggtgctgagcagcaaatcctggtgagcaaccacgcggatggtctcctgcgtggg 554
 DB 216 cccgtctcagcagaagaattcctcgcagaatcgcggacatgctcctcgcgcgg 275
 QY 555 catgttctctgctgggactatgttcgagggcagcggcgcgagatgctgacgtgttcc 614
 DB 276 gatgttctctcgtgggctgtgttcgaggggaacagcagcagcaccatcctgtttct 335
 QY 615 caccctgcagcagtagctgttcc---cagcgggaagggtaccctcgggtccaggac 671
 DB 336 cactctcagcagcagtggtgtcctcgcagcagaggaacaagccagcggctcaagtc 395
 QY 672 cctttaccattatgggtcaaaagatctgcccacagtggttctctacatgctgtggccat 731
 DB 396 cctctattatggtgtcaaaagatttggccacgggtttctctacatgctgtggccat 455
 QY 732 catctatcagccaccattcagagtagctgtgtagataagctcagcgggagactgcagct 791
 DB 456 cattattcgcacaaattcaggaatgtgttggaataaattcaacaagaatgcagtt 515
 QY 792 caccagaagcaaaacaaatgaatgagcggcggcagcagcagtagtgttctacatagt 851
 DB 516 caccagaagcaaaacaaagtttaacagtagtgcgtgctgtagtgttctactttt 575
 QY 852 gtctgtatctgggtgatgatcttgcctctgagaaactgcctgcagacccacctct 911
 DB 576 ttctgtatttggggcacattcatttaactctctgaaaactgcctgcagacccacctct 635
 QY 912 attgtggaagtctcagccacacacatgatgacatttcagatgaattttctacatctc 971
 DB 636 tatatgggaagcctgcctccatagatgatgacatttcaaatgaagttttctacgtatc 695
 QY 972 acagttggttactggtttcagtttcccgagagctctacttccagaaagtcaggaaaca 1031
 DB 696 ccagttggttactggtttcagcttctcctgaactctacttccagaaacccaaacaa 755

QY 1032 agatcccggtgtcaactcatctacatgtgctccacactctccacattggaggggccta 1091
 DB 756 agacatccctcgtcaactgtctacatgtggtctccacactctccacattactggagctta 815
 QY 1092 tctctgtacttgaaacacacgtggtcgtctgtctgtatgtgcactatgtctgcagct 1151
 DB 816 tctctgtacttgaaatcatttgggactctctctcttctgttactgcattatttggtaatt 875
 QY 1152 cctctcagcgtgtgcagcctcgttacttctgtgggatgagcgggtaccagaaaggtgtgc 1211
 DB 876 acttccacatgtcggcctgttacttctgtgtgaaagtcacagaaagcagcatc 935
 QY 1212 ttgtggcctatgtgttatatccgggagactgcagactgcacactgattctcagttgttac 1271
 DB 936 tctgtggcctatgtgttatctctgtgtgtagactgtgactttaattgttccgtactcac 995
 QY 1272 agtagggtcctcactggcgg---gacaaatcggaatggaatgctctctctctgttaagt 1328
 DB 996 tgtgggttccactggcgtggtgcagatcggaatcgtgactgactgacttactggaatgt 1055
 QY 1329 caatgttggcagcctaaatcgtctgtctcctcagtgtagtgcagtatccaggtgtacat 1388
 DB 1056 aaatgttggcagcctaaatgtctgtctcctcagtgtagtgcagtatccagcctcag 1115
 QY 1389 aacatgacctgcagcctgtcgtcagagatggttagaagatgcgaatcttcatgt 1448
 DB 1116 aacatggaacttaattactctcgtcagaggtggtgtagaagattcttaattcagc 1175
 QY 1449 ctgtgggaggaa-----gagacggtccagagaaagacacagaaagcagaaatgg 1496
 DB 1176 ctcattgatgaaagaaacgggtcagatctcttaaaaaagacagaaacggagtg 1235
 QY 1497 agtggagaatcccaatagaatagattctccacccaaagaaagaaagacagccttagca 1556
 DB 1236 agtggaaactcaaatagtagtagactgtccccaagagaaagaaatcttcataatc 1295
 QY 1557 gttcgaagcgaattgattcttaccctcgaaggaaatccactctcttcttattgtgtctct 1616
 DB 1296 ttgcgaagcgaattgatttaattgtctgcaaaagaaat---ctgctcttggaggttctct 1352
 QY 1617 gtgcagaga-tttctgttctcagaaacgggtgctgtctttttgaaatattgctaattgtat 1675
 DB 1353 gcactagagattttctgttttgaaatagttcgtcgtctctgtttttgtattgaac 1412
 QY 1676 tgtctaattgtttt-----aaagtttgcagcagtagtgcgtgg 1716
 DB 1413 tgttctatgtattttttaaagacatttgaggggaggtattattatgaatgggaaaaa 1472
 QY 1717 ggatgggggttaagactaaacacactcagcctcctaaatcacagatagtttaacgg--- 1773
 DB 1473 agatttgggtgagactaaattactcatcgtcgaataatgtcaaaatagttttgggat 1532
 QY 1774 -accaacatttattgttaggtttcttaccctcagcagatttcccaaacggtttgtggtga 1832
 DB 1533 caccactattttgtttgatttttaaccttcaacatttccctaatgatttgagaga 1592
 QY 1833 tgactgcagaattgtacataataatagt 1863
 DB 1593 taactgcacaatttgcataatgatact 1623

RESULT 6

AAF74782

ID AAF74782 standard; cDNA; 1288 BP.

XX AAF74782;

AC AC

DT 17-MAY-2001 (first entry)

XX Human WAR-1 encoding cDNA sequence SEQ ID NO:5.

DE WAR-1; protein screening; endoplasmic reticulum membrane protein;

XX endoplasmic reticulum membrane transport; secretory protein;

KW


```

ID AAF74780 standard; cDNA; 1110 BP.
XX AC AAF74780;
XX DE
XX DT 17-MAY-2001 (first entry)
XX XX Human WAR-1 encoding cDNA sequence SEQ ID NO:3.
XX KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
XX KW endoplasmic reticulum membrane transport; secretory protein;
XX KW cell membrane protein; cytosolic; CNS active; antiallergic; cancer;
XX KW antirheumatic; nervous system disorder; immune disorder; allergy;
XX KW rheumatism; skeletal disorder; ss.
XX OS Homo sapiens.
XX FH
XX FT Key Location/Qualifiers
XX FT CDS 1..1110
XX FT /*tag= a
XX FT /product= "WAR-1"
XX WO200114582-A1.
XX PD 01-MAR-2001.
XX XX 17-AUG-2000; 2000WO-JP05488.
XX PF
XX XX 20-AUG-1999; 99JP-0234764.
XX PR
XX PA (SUMU ) SUMITOMO PHARM CO LTD.
XX PI Tohdoh N, Okuyama H, Imamura M, Ishikawa H, Nemoto K;
XX XX
XX DR WPT: 2001-202940/20.
XX DR P-PSDB: AAB70695.
XX PT Transformation of a cell with separate vectors expressing the sense and
XX PT antisense strands of WAR-1 DNA for screening secretory and membrane
XX PT proteins expressed by the cell -
XX PS Claim 2: Page 62-65; 79pp; Japanese.
XX XX
XX CC The present invention describes a screening method for secretory and
XX CC membrane proteins consisting of transformation of a cell with separate
XX CC expression vectors for the sense and antisense RNA of DNA encoding an
XX CC endoplasmic reticulum membrane protein participating in endoplasmic
XX CC reticulum transport of proteins. Also described are: (1) secretory and
XX CC cell membrane proteins identified by the screening method; (2) drug
XX CC compositions containing these proteins; (3) host cells transformed by
XX CC the separate expression vectors of the method; and (4) the preparation
XX CC of secretory and cell membrane proteins by culture of the transformants.
XX CC The method can be used for the identification and preparation of
XX CC proteins for use in the treatment and prevention of diseases such as
XX CC cancer, disorders of the nervous system, immune disorders (including
XX CC allergies and rheumatism) and skeletal disorders. The present sequence
XX CC encodes a specifically claimed human WAR-1 protein from the present
XX CC invention.
XX SQ Sequence 1110 BP; 296 A; 241 C; 236 G; 337 T; 0 other;

Query Match 28.0%; Score 647.8; DB 22; Length 1110;
Best Local Similarity 76.1%; Pred. No. 2.7e-116;
Matches 844; Conservative 0; Mismatches 247; Indels 18; Gaps 3;

.QY 463 atgggggtccgcaagaagcgcaggaaccccccggtgtgtgtgagccacgaattcatgtgtg 522
DB 1 atgggggtccgtaagaagagcaccagaaccccccggtgtgtgtgtgtgtgtgtgtgtgtgt 60
QY 523 cagaacacacggtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 582
DB 61 cagaatcatcgacatcgtctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120

```



```
Db 1185 gcagatcttctaaaaaggaacagaaatggtgtgaatggagaaacatttaacttcaaatgtag 1244
QY 1517 tagattctcaccacaagaagaaagc-tccttagcagttgcaagcaattgtattc 1575
Db 1245 cagactctcccccgaatacaaaagagaaatctctcaatgaattataaactaatgtatta 1304
QY 1576 ttaccctccaaaggaatccactctctcttattgtgtgtctctgtgctagagattttctgtt 1635
Db 1305 atgtccccaagaatactgctcttactatactcttcagcattagagattttctgttct 1364
QY 1636 ctccagaagcggctgctgttttgaatattgctaattgttctataatgtgttttaagg 1695
Db 1365 tgaataacagctgtgctcttattgtattgtattgtacggtttcactttt----- 1419
QY 1696 ttttcagacgtatgattggggaatgggggttaagactaaacacactcagct----- 1747
Db 1420 ttttaagggcatttgggggagattattgctatgaatgaataaaacatttttagcttaga 1479
QY 1748 -----ctaaatcacgtcagaatgttaacggaacacacattcttatttagtttcttac 1802
Db 1480 ctaagctacctgcttcaaaatagtttagggaccacacacatttttatttttttat 1539
QY 1803 ctca-acgattttccaaacgttttgggtgtagctg-----cagaatttgttacataa 1856
Db 1540 ttttgaaacatttttcaatgatttggagagaaactatttacaaaaattccacatcag 1599
QY 1857 taatag--tttctcgtctccaatgttcttcttatcgaattaaacaagtctgctagcaagtgg 1914
Db 1600 tgatacaatttctgtgcacaaatttttataatagcagagtggtgctgttctcaagaag 1659
QY 1915 ttgttttctcaatgttctcctcaggaataaagtggaaaaatc-----tgataa 1962
Db 1660 gccataatttttaagtattcttcagggttaacatggaaataactataaagttagtgc 1719
QY 1963 aggttaaacccaatacagtaattatgtaaacggttgggaatttttttaaagtggttttaattt 2022
Db 1720 actttaatatttttcagtgctctcaatttttttgggaattttgt-agaactttacacctg 1778
QY 2023 acaatggaaagcatttgcacacacacacaaataatgtgttaatttttatgtagtaatt 2082
Db 1779 gaaaaaagatttgaataaccccggaacaaattgtgtcttatttttataggtagtggtt 1838
QY 2083 gttagtgttacgccccccataaagcatcaataataga 2120
Db 1839 attagttattacatccccatttttaaaaaacaaaacataa 1876

RESULT 9
AAS62269
ID AAS62269 standard; cDNA; 2756 BP.
XX
AC AAS62269;
XX
DT 14-FEB-2002 (first entry)
XX
DE cDNA sequence #56 encoding novel human secreted protein.
XX
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; gene therapy; antimicrobial; hepatotropic;
KW immunosuppressive; antirheumatic; ss.
XX
OS Homo sapiens.
XX
PN WO200177291-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US10485.
XX
PR 06-APR-2000; 2000US-195604P.
XX
PA (GEMY ) GENETICS INST INC.
```

```
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
XX WPT; 2002-010900/01.
XX
PT New polynucleotides encoding secreted proteins useful for treating e.g.
PT asthma, HIV and Crohn's disease -
XX
PS Claim 1; Page 103-104; 391pp; English.
XX
CC The present invention relates to the isolation of novel cDNA sequences
CC which encode human secreted proteins. The cDNA sequences have been
CC derived from a variety of human tissues. The invention also provides
CC a method for producing proteins from these polynucleotide sequences.
CC The proteins are useful for identifying compounds that modulate their
CC activity and production, and the cell is also useful for identifying
CC compounds that modulate expression of the polynucleotide sequences
CC encoding the secreted proteins. The sequences of the invention are
CC useful for treating diseases such as hyperproliferative disorders
CC (e.g. cancer), immune deficiency disorders (e.g. severe combined
CC immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory
CC disorders (e.g. arthritis) and infectious disorders (e.g. hepatitis).
CC The polynucleotide sequences of the invention are also useful in gene
CC therapy. AAS62214-AAS62838 represent the cDNA sequences of the
CC invention that encode for novel human secreted proteins.
XX
SQ Sequence 2756 BP; 825 A; 498 C; 505 G; 928 T; 0 other;

Query Match 22.4%; Score 517.8; DB 24; Length 2756;
Best Local Similarity 60.5%; Pred. No. 5,1e-91;
Matches 1113; Conservative 0; Mismatches 652; Indels 74; Gaps 13;

QY 350 gtcggcagcgtgcgaccctggcgccagacggcggggatggggagcccgctggg 409
Db 31 ggcctgcagcggggcgctgagggagcagccagcggggcggtcggtgagcagct 90
QY 410 agcggcgagtgatcagcgggtggcgccggtgagtagcggtagtaccggtgaggggc 469
Db 91 gggagagagacaacggggcgagcacctgcagggcgggcgccgccacacattgcca 150
QY 470 tccgcaagaagaacccaggaaccccggtgctgagccacgaattcattggtgcagaacc 529
Db 151 ttcgcaagaagaacccaggaaccccggtgctgagccacgaattcattggtgcagaacc 210
QY 530 acgcggtatggtctcctgcgtggcgatgtcttcgtgctgggacttatgttcgagggca 589
Db 211 acgcggtatggtctcctgcgtggcgatgtcttcgtgctgggacttatgttcgagggca 270
QY 590 cggcgagatgctgctgcttcctccacccgtcagcagcagcagcagcagcagcagcag 646
Db 271 cggcaaaagcttctatcttcttcttcttcttcttcttcttcttcttcttcttcttct 330
QY 647 aagggtaccctcgggggtccaggaccctttaccattatgggttcaagatctggccacag 706
Db 331 aagaacaagctactgaatcagtcctcttattacttggtcgcacaaagatttgggtactg 390
QY 707 tgtcttctacatgctggtggccatcattcaccgcccacacattcagagtagtgcgtag 766
Db 391 tttcttctacatgctggtggccatcattcaccgcccacacattcagagtagtgcgtag 450
QY 767 ataagctcagcggagactcagctcaccagaaggcaaaacaaacaaatgaatgaggccg 826
Db 451 ataaaattaacaggcgaatgcattctccaaacaaacaaacacacagcaagttaataatcg 510
QY 827 ggcagctgagtgcttctacatagtgctggtatctggtgggtatgatcattctggcctctg 886
Db 511 gtcagcttagtgcttctacatttttgcctgtgttggggcaccattcattctctctctg 570
QY 887 agaactgctgcagacccactctatgttggaagcttcagcccccacacatgatgacat 946
```

Db 571 aaaacacatctcagaccacaactattttatgaggaggttatccccataacacctgatgacat 630
Qy 947 ttcagatgaattttttacatctcacagttggtctactggtttcatagttttcccgagac 1006
Db 631 ttcacaaatgaagttttttacatatacacagctggtctactggtttcatgcttttctctgaac 590
Qy 1007 tctacttcagaaaatcaggaaacagatataccccgggtcaactcatctacattggcctcc 1066
Db 691 tctacttcagaaaacccaaaagaagatattcttcctgcgtcoagctgtctacattggtcttt 750
Qy 1067 acctttccacattggaggggtcctatctcttgaattgaacacacotggcctgctgcttc 1126
Db 751 acctttccacattgctggaggttaccctttgaacttgaatcatctagagactgttcttc 810
Qy 1127 tgatgctcactatgctgctgagctccctccagcgtgtgcagcctgcttactttgggg 1186
Db 811 tgggtctacattatttttgggaattttcttccacatttcccgctgtttttatttttagca 870
Qy 1187 atgagcgttaccagaaaaggtttgtttgtggtccctatcgttttatatccgggagactcg 1246
Db 871 atgaaaagtatacagaaggattttctctgtggcgagttcttttgttttgggaagacttc 930
Qy 1247 tga-cactgattgtctagtggttacagtagggttcacttggccgggac---aaatcgg 1302
Db 931 tgcacttcaattcttctagctagctgttgggttgggcttgcagagcagaaaatcag 990
Qy 1303 aatgaaaatgctctctctggttaagtgtcaatgtgttggcagctaaatcgtctgtctgc 1362
Db 991 aagctggtatttcagctactggaaactcaatgtgttagctgttagaatcgtcttctgcga 1050
Qy 1363 tgcagttcagtatccaggtgtacatacatggaacctgtgacacgctgtggtctcagaga 1422
Db 1051 tccatttgcgttactcaggcatttatgatgtggaagtccaatttttccagcttcgagag 1110
Qy 1423 tgggttagaagatcgcaattcttcatctgtgtgggaggaagagagcgtccagg----- 1473
Db 1111 tggagggaacattctgcttttcaggccaccgctgtgagaagaacccaacagtaactaaa 1170
Qy 1474 -----tcgagaaaagccacagaaaatggagtgga-----aatccaaataga 1515
Db 1171 ggcagatcttctaaaaaggacagaaaatggtggaatggaaacttaactcaaatgta 1230
Qy 1516 atagatttccaccaaaagagaaaagagaaagc-tccttagcaggttgcagcgaattgatt 1574
Db 1231 gcagactctcccggaataaaaaagagaaaatcttcataatgaattataaaactaattgatt 1290
Qy 1575 cttaactccaagggaatccactctcttcttattgtgtctgtctgtctagagattttctgt 1634
Db 1291 aatgtccccaagaatactgcttctactatcttccagcattagagagatttttctgttc 1350
Qy 1635 tcttcagaaacgggtcgtgctttttgaaatattgctaatgtattgtctaatgttttttaag 1694
Db 1351 ttgaaaatacagctgtgctctgtatttttgcattgttgcaggttttcagctattt---- 1406
Qy 1695 gttttgcagacgtatgagtgggggtatggggttagactaaaccactcagcct----- 1747
Db 1407 -tttaaaaggcatttgagggggagattattgctatgaatgaaaaaatttttagcttag 1465
Qy 1748 -----ctaaatacagctcagaatgttaacggacccaacatctatttttagttgcttta 1801
Db 1466 actaagctaccgcttcaaaaatagtttagggaccaccacacataattttttgttttta 1525
Qy 1802 cctca-acgattttccaaacgttttgggtgatgactg-----cagaattgtgtacataa 1855
Db 1526 tttttgaacatttttctaatgatttggagagaaaactatttcaaaaatttccacataca 1585
Qy 1856 ataataag---tttctgcttccaaatgttctttatcgaaataacagctgctgtagcaagtg 1913
Db 1586 gtgatacaattttctgtctcaacaaattttttataatgacagagtggtctgtcttaagaa 1645
Qy 1914 gttgtttttctcaatgttctctcagagataaagtggaatac-----tgata 1961
Db 1646 ggcacataatttttaagttatcttccagggttaacatgggaataactataaagttggatgta 1705

Qy 1962 aaggtttaaactcaaatcagctatttatgaaccgttgggatttttttaaagtgttttaaat 2021
Db 1706 aactttaaatgttttcagtggtctctctaatttttttgaattttgt-agactttacacct 1764
Qy 2022 tacaatggaagcatttgcataaacaccacaaaataatgtgttttaatttatgagtagta 2081
Db 1765 ggaataaaagatttgaatacaccggaacaattgtgtgtttttattataggtagtggt 1824
Qy 2082 tgttaagtgttaacgccccattaaagcatcaaaataga 2120
Db 1825 tattagattatacctccccatttttaaaacaaaacataa 1863

RESULT 10
AAS18576
ID AAS18576 standard; cDNA; 1517 BP.
XX
AC AAS18576;
XX AC
DT 12-MAR-2002 (first entry)
XX DE cDNA encoding human translocating chain-associated membrane protein.
XX KW Human; translocating chain-associated membrane protein; BiotRAM; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 153..790
FT FT /*tag= a
FT FT /product= "Translocating chain-associated membrane
XX PN CN1310184-A.
XX PD 29-AUG-2001.
XX PF 24-FEB-2000; 2000CN-0111729.
XX PR 24-FEB-2000; 2000CN-0111729.
XX PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.
XX PI Mao Y, Xie Y;
XX WPI: 2002-034947/05.
XX P-PSDB; AAU10977.
XX PT New human transposition chain related membrane protein and its coding
XX PT sequence -
XX PS Claim 6; Page 18; 22pp; Chinese.
XX CC The invention relates to a novel human translocating chain associating
XX CC membrane protein (BiotRAM), polynucleotides encoding this polypeptide
XX CC and the recombination process used to produce the polypeptide. The
XX CC present invention also discloses the method of applying the polypeptide
XX CC and polynucleotides in treating immunological disorder, malignant tumour,
XX CC cancer and other diseases. The antagonist resisting the polypeptide and
XX CC its treatment effect is also disclosed. Diagnosis and determination
XX CC method based on the discrimination of the mutation in the nucleic acid
XX CC sequence and the change in the polypeptide expression level, and the
XX CC application of the polynucleotides encoding the BiotRAM. The present
XX CC sequence represents the coding sequence of the novel human BiotRAM
XX CC protein as described in the invention.
XX SQ Sequence 1517 BP; 459 A; 262 C; 308 G; 488 T; 0 other;

Query Match 20.7%; Score 477.6; DB 24; Length 1517;
Best Local Similarity 64.6%; Pred. No. 2.7e-83;
Matches 913; Conservative 0; Mismatches 439; Indels 62; Gaps 11;

QY 555 catgttctctgtggtgacttatgttcgagggcagcgccgagatgctcgatcgtgttctt 614
 Db 276 gatgttctctgtggtgacttatgttcgagggcagcgccgagatcctcctgttctt 335
 QY 615 caccctgcagcagtgagctgttcttcc---cagcggaagggtaccctcggggtccaggac 671
 Db 336 cactcttcagcagtgagctgttcttccctgcagcagaggaacacccagcggtcagaagtc 395
 QY 672 cctttaccattatgggttcacaaatcgtccacagtgcttctcagatgctgtggccat 731
 Db 396 cctctattatattggttcacaaatcgtccacagtgcttctcagatgctgtggccat 455
 QY 732 catcattcagcagcaccattcagcagtgctgtagataagctcagcgagactgcagct 791
 Db 456 cattattcagcagcaccattcagcagtgctgtagataagctcagcgagactgcagct 515
 QY 792 caccaaaggcacaacaaataatgaatgagcgccgagcggtgagtggttctacatagt 851
 Db 516 caccaaaggcacaacaaataatgaatgagcgcggtgagtggttctacatagt 575
 QY 852 gtctgtgattcgtggttatgatcattcgtcctcagaaactgctcagaaactcctc 911
 Db 576 ttctgtatttggggacattcatttctcgtgaaactgctcagaaactcctc 635
 QY 912 attgtggaagtcagcagcaccacacatgatgacatttcagatgaatttttctacatctc 971
 Db 636 tatatggaagtcagcagcaccacacatgatgacatttcagatgaatttttctacatctc 695
 QY 972 ac-agttggcttactggttctacatgcttcccgagctcacttcacagaaagtcaggaaac 1030
 Db 696 ccaatttggcttactggttctacatgcttcccgagctcacttcacagaaagtcaggaaac 755
 QY 1031 aagatccccgggtcaactcatcattcagcagcaccctcctccaca 1078
 Db 756 aagacatnccctgt-aacttggctacattggncttcaactttttcaca 802

RESULT 13

AAK93837

ID AAK93837 standard; cDNA; 802 BP.

XX AC AAK93837;

XX DT 06-NOV-2001 (first entry)

XX DE Human cDNA clone representative sequence, SEQ ID NO: 2297.

XX KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX OS Homo sapiens.

XX PN EP1130094-A2.

XX PD 05-SEP-2001.

XX PF 07-JUL-2000; 2000BP-0114089.

XX PR 08-JUL-1999; 99JP-0194486.

XX PR 11-JAN-2000; 2000JP-0118774.

XX PR 02-MAY-2000; 2000JP-0183765.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX DR WPI; 2001-524255/58.

XX PT 830 Primers useful for synthesizing full length cDNA clones and their

XX PT use in genetic manipulation -

XX PS Example 11; SEQ ID NO 2297; 1380pp + sequence listing; English.

XX XX

CC The invention relates to primers for synthesizing full length cDNA
 CC clones, 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence was used as the
 CC representative sequence from a human clone which was used in
 CC homology searches to identify the clone.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 SQ Sequence 802 BP; 186 A; 200 C; 193 G; 220 T; 3 other;

Query Match

16.1%; Score 372.2; DB 22; Length 802;

Best Local Similarity 77.3%; Pred. No. 6.2e-63;

Matches 501; Conservative 0; Mismatches 141; Indels 6; Gaps 4;

QY 435 gccggtgagtagcggtagtaccgcggtacggggtcgcgaagaagacgcaggaaccc 494

Db 157 gccggtgagtagcggga-agcgccaccatggggtcgcgaagaagacgcaggaaccc 215

QY 495 cccggtgctgagccacgaattcattggtcagaacacacgcggatattggtctcgtggtg 554

Db 216 cccggttctcagcagggaattcattggtcagaacacacgcggatattggtctcgtggtg 275

QY 555 catgttctctgtcgtggaattatgttcgagggcagcgccgagatgctcgatcgtgttct 614

Db 276 gatgttctctgtcgtggaattatgttcgagggcagcgccgagatcctcgtgttct 335

QY 615 caccctgcagcagtgagtcgttctcc---cagcggaagggtaccctcggggtccaggac 671

Db 336 cactctcagcagtggtgtcgttcctgcagcagaggaacacgcagcggtcctcaagtc 395

QY 672 cctttaccattatgggtcacaagatcgtgccacagtgcttctctacatgctgtgacct 731

Db 396 cctctattattatgggtcacaagatcgtgccacagtgcttctctacatgctgtgacct 455

QY 732 catcattcagcagcaccattcagcagtgctgtagataagctcagcgagactgcagct 791

Db 456 cattattcagcagcaccattcagcagtgctgtagataagctcagcgagactgcagct 515

QY 792 caccaaaggcacaacaaataatgaatgagcgccgagctgagtgcttctacatagt. 851

Db 516 caccaaaggcacaacaaataatgaatgagcgctgctggtcagtttagtgttctactttt 575

QY 852 gtctgttatctggtgtagtattcattcgtcctcagaaactcctcgtcagacccactct 911

Db 576 ttctgtatttggggcagcattcattttaaactcctgaaactcgtcagacccactct 635

QY 912 attgtggaagtcagcagcaccacacatgatgacatttcagatgaatttttctacatctc 971

Db 636 tatatggaagtcagcagcaccacacatgatgacatttcagatgaatttttctacatctc 695

QY 972 ac-agttggcttactggttctacatgcttcccgagctcacttcacagaaagtcaggaaac 1030

Db 696 ccaatttggcttactggttctacatgcttcccgagctcacttcacagaaagtcaggaaac 755

QY 1031 aagatccccgggtcaactcatcattggtcctcctcactcttccaca 1078

Db 756 aagacatnccctgt-aacttggctacattggncttcaactttttcaca 802

RESULT 14

AAS47389/C

ID AAS47389 standard; cDNA; 642 BP.

XX AC AAS47389;

XX DT 18-DEC-2001 (first entry)

Query Match		14.98;	Score 345;	DB 22;	Length 642;
Best Local Similarity		72.0%;	Pred. No. 1.le-57;		
Matches 450;		Conservative 0;	Mismatches 175;	Indels 0;	Gaps 0;
Qy	641	cagcgggaagggtaccctcggggtccaggaccctttaccattatgggggtcaaaagatctgg	700		
Db	625	CAACAGAGAAGACAGCTACTGAATCAGTGTCCCTTTATTACTATGGCAATCAAGATTGG	566		
Qy	701	ccacagtgttctttacatgctggtggccatcattcaacgccaccattcaggagtacg	760		
Db	565	CTACTGTTTTCTTCTACATGCTAGTGGCGATAATATTATCATGCCGTAATTCGAAGAGTATA	506		
Qy	761	tgctagataagctcagccggagagctcagctcaccaaaggcaaaacaaataatgaatg	820		
Db	505	TGTTGGATAAAATTAACAGGGGAATGCATCTCCAAAACAAAACACAGCAAGTTTAATG	446		
Qy	821	aggcgggagcagctgagtgtgttctacatagtgctggttatctggggtatgatcattctgg	880		
Db	445	AATCTGGTCAGCTTAGTGCCTTACCTTTTGGCTGTGTTGGGGCACATTCATCTCA	386		
Qy	881	cctctgagaactgctgctcagacccccactctattgtggaagtctcagccccacaacatga	940		
Db	385	TCTCTGAAACTACATCTCAGACCCAACTATCTTATGAGGGCTTATCCCATTAACCTGA	326		
Qy	941	tgacatttcagatgaatttttctacatctcacagttgcttactggttttcattagtttcc	1000		
Db	325	TGACATTTCAAAAGTTTTTCACATATCACAGCTGGCTTACTTGGCTTCATGCTTTTC	266		
Qy	1001	cggagctctacttccagaaaagtcaggaacaagatatcccggttcaactcatctacattg	1060		
Db	265	CTGAACCTACTTCCAGAAACCAAAAGAGATATTCCTCGTCAGCTTGTCTACATTG	206		
Qy	1061	gcctccaccttccacattggaggggccctatctctgttaattgaaccacctgggacctgc	1120		
Db	205	GTCCTTACCTCTCCACATTCCTGGAGCTTACCTTTTGAACCTGAATCATCTAGGACTTG	146		
Qy	1121	tgetctgatgctgactatgctgctcagctccctcctcagcgtgctgagcctgctttact	1180		
Db	145	TTCTTCTGGTGTACATTAATTTTGTGAAATTCCTTTCCACATTTCCCGCTGTITTAAT	86		
Qy	1181	ttgggggatgagcgggtaccagaaaagggttgtcttcttggccctatcgtgtttatatccggga	1240		
Db	85	TTAGCAATGAAAGTATCAGAAAGGATTTTCTCTGTGGGCAGTTCTTTTGTGTTGGGAA	26		
Qy	1241	gactcgtgacactgattgtctcagt	1265		
Db	25	GACTTCTGACTTTAATTCTTTTCAGT	1		

Search completed: September 6, 2002, 14:57:08
Job time: 9842 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 13:08:16 ; Search time 97.09 Seconds
(without alignments)
5846.730 Million cell updates/sec

Title: US-09-807-470-1

Perfect score: 2311

Sequence: 1 agagagagagagagagagag.....aaaaaaaaaaaaaaaaaaaaa 2311

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67	2.9	958	2	US-08-757-046A-5
2	67	2.9	958	3	US-09-447-208-5
3	67	2.9	958	3	US-09-135-988-5
4	67	2.9	958	4	US-09-277-716-5
5	67	2.9	958	4	US-08-597-274A-5
6	64.2	2.8	1474	4	US-08-821-994-64
7	64	2.8	1582	3	US-08-545-196B-10
8	64	2.8	1582	3	US-08-545-196B-10
9	63.4	2.7	2674	4	US-09-817-180-1
10	63.4	2.7	2852	3	US-09-027-137-2
11	63.2	2.7	1441	4	US-08-821-994-63
12	63	2.7	1315	4	US-09-721-822A-10
13	63	2.7	1512	2	US-08-909-965C-8
14	63	2.7	5173	1	US-08-242-677-1
15	62.8	2.7	2447	2	US-09-014-969-14
16	61.4	2.7	1454	4	US-09-372-422A-19
17	60.8	2.6	1882	4	US-09-370-253-1
18	60.6	2.6	2862	1	US-08-148-209A-1
19	60.6	2.6	2989	6	5378464-1
20	60.4	2.6	1364	1	US-08-265-087-3
21	60.4	2.6	1364	1	US-08-621-493-3
22	60.4	2.6	1364	2	US-08-965-688-3
23	60.4	2.6	1364	4	US-09-260-173-3
24	60.2	2.6	144	1	US-08-702-344-26
25	60.2	2.6	240	1	US-08-628-417-6
26	60.2	2.6	1066	1	US-08-157-101A-4
27	60.2	2.6	1736	3	US-09-182-816-22

c 28 60.2 2.6 1736 3 US-09-182-816-24 Sequence 24, Appl
29 60.2 2.6 1736 3 US-09-471-528-22 Sequence 22, Appl
c 30 60.2 2.6 1736 3 US-09-471-528-22 Sequence 24, Appl
31 60.2 2.6 1736 4 US-09-634-530-22 Sequence 22, Appl
c 32 60.2 2.6 1736 4 US-09-634-530-24 Sequence 7, Appl
33 60.2 2.6 2407 4 US-09-370-807-7 Sequence 24, Appl
34 60 2.6 536 1 US-08-341-568-1 Sequence 1, Appl
35 60 2.6 536 2 US-08-911-020-1 Sequence 1, Appl
36 60 2.6 1493 1 US-08-340-820-24 Sequence 24, Appl
37 60 2.6 1493 1 US-08-593-535-24 Sequence 24, Appl
38 60 2.6 3080 4 US-09-099-041A-25 Sequence 25, Appl
39 59.8 2.6 2671 6 5168051-9 Patent No. 5168051
40 59.6 2.6 1172 1 US-07-945-288-9 Sequence 9, Appl
41 59.6 2.6 1172 1 US-08-462-831-9 Sequence 9, Appl
42 59.6 2.6 1172 1 US-08-461-809-9 Sequence 9, Appl
43 59.6 2.6 1172 1 US-08-461-441-9 Sequence 9, Appl
44 59.6 2.6 1172 5 PCT-US93-08518-9 Sequence 9, Appl
45 59.6 2.6 1198 3 US-09-248-335-27 Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-08-757-046A-5
; Sequence 5, Application US/08757046A
; Patent No. 5876995
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; TITLE OF INVENTION: BIOLUMINESCENT ARTICLES OF MANUFACTURE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,046A
; FILING DATE: 11-25-96
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,274
; FILING DATE: 02-06-96
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6680-105B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 115..702
; OTHER INFORMATION: apoaequorin-encoding gene

```
/ PUBLICATION INFORMATION:
/   AUTHORS: Inouye et al.
/   JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
/   VOLUME: 82
/   PAGES: 3154-3158
/   DATE: (1985)
/   DOCUMENT NUMBER: PATENT NO.: 5,093,240
/   US-08-757-046A-5

Query Match          2.9%; Score 67; DB 2; Length 958;
Best Local Similarity 63.2%; Pred. No. 4.6e-07;
Matches 103; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 2149 tttagcgaaatcaagatacctttaataatggtggttactaaagaagtaaacgacttc 2208
Db 792 TTGAACGATTCATCGTTGCTGTTGATTTTGTATATTAGGAACAGATTAAATCGAATGA 851
QY 2209 ttctgtttattttaaacactgttacaggaaactcgcaaaataataattactgaaaa 2268
Db 852 TTAGTTGTTTATCAACAGAACTTACAAATCGAAAAAGTAAAAAAGTAAAAAAGTAAAAA 911
QY 2269 aaaaaaagtaaaagtaaaagtaaaagtaaaagtaaaagtaaaagtaaaagtaaaag 2311
Db 912 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 954

RESULT 2
US-09-447-208-5
; Sequence 5, Application US/09447208
; Patent No. 6113886
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; TITLE OF INVENTION: BIOLUMINESCENT ARTICLES OF MANUFACTURE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/447,208
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 0909/135,988
; FILING DATE: 08-17-98
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/757,046
; FILING DATE: 11-25-96
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,274
; FILING DATE: 02-06-96
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24727-105C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-450-8400
; TELEFAX: 619-450-8499
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 958 base pairs
```

```
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE:
/ ORIGINAL SOURCE:
/ FEATURE:
/ NAME/KEY: Coding Sequence
/ LOCATION: 115...702
/ OTHER INFORMATION: apoaequorin-encoding gene
/ PUBLICATION INFORMATION: PATENT NO.: 5,093,240
/ PUBLICATION INFORMATION:
/ AUTHORS: Inouye et al.
/ JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
/ VOLUME: 82
/ PAGES: 3154-3158
/ DATE: (1985)
/ US-09-447-208-5

Query Match          2.9%; Score 67; DB 3; Length 958;
Best Local Similarity 63.2%; Pred. No. 4.6e-07;
Matches 103; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 2149 tttagcgaaatcaagatacctttaataatggtggttactaaagaagtaaacgacttc 2208
Db 792 TTGAACGATTCATCGTTGCTGTTGATTTTGTATATTAGGAACAGATTAAATCGAATGA 851
QY 2209 ttctgtttattttaaacactgttacaggaaactcgcaaaataataattactgaaaa 2268
Db 852 TTAGTTGTTTATCAACAGAACTTACAAATCGAAAAAGTAAAAAAGTAAAAAAGTAAAAA 911
QY 2269 aaaaaaagtaaaagtaaaagtaaaagtaaaagtaaaagtaaaagtaaaagtaaaag 2311
Db 912 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 954

RESULT 3
US-09-135-988-5
; Sequence 5, Application US/09135988
; Patent No. 6152358
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; TITLE OF INVENTION: BIOLUMINESCENT ARTICLES OF MANUFACTURE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,988
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/757,046
; FILING DATE: 11-25-96
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,274
; FILING DATE: 02-06-96
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24727-105C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-450-8400
; TELEFAX: 619-450-8499
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 958 base pairs
```



```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 115...702
; OTHER INFORMATION: apoaequorin-encoding gene
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: 5,093,240
; AUTHORS: Inouye et al.
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 82
; PAGES: 3154-3158
; DATE: (1985)
US-08-597-274A-5

Query Match      2.9%; Score 67; DB 4; Length 958;
Best Local Similarity 63.2%; Pred. No. 4.6e-07;
Matches 103; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 2149 tttagcgcaatcaagatacctttaataatgtgtgggtactaaagaagtaaacgacttc 2208
Db 792 TTGAACGATTCAATCGTTTGTGTGTTGTTTGTAAATTAGGAACAGATTAAATCGAATGA 851

Qy 2209 ttctgtttatttaaacactgtacaggaaaactcgcaaaataaataattactgaaaaa 2268
Db 852 TTAGTGTGTTTTTAAATCAACAGAACTTACAAATCGAANAAGTAAATAAAAAAAAAA 911

Qy 2269 aaaaaaiaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2311
Db 912 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 954

RESULT 6
US-08-821-994-64
; Sequence 64, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 1474
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-64

Query Match      2.8%; Score 64.2; DB 4; Length 1474;
Best Local Similarity 49.0%; Pred. No. 2.8e-06;
Matches 171; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

Qy 1963 aggttaactcaaatcagttattgtaacccgttggttttttaaaagtttttaaaattt 2022
Db 1116 aagtcacagtcacaaacccgttcgtggaaccagcagtggttgaaagtttttaaaataaaact 1175

Qy 2023 acaatgaaagcatttgtcaaacaccacaaataatgtgttttaattttatagtagtaatt 2092
Db 1176 caattgggagttttataactaaagatttaattctcatattattgtttgtatgtatgtat 1235
```

```

Qy 2083 gttagtgttacgccccattaaagcatcaaaaaatgatgaatagatgacatgtgtgtgata 2142
Db 1236 caaaaaagaaggtatttgatccagcagcaggtatttagctgtataaatcccttatgtcgat 1295

Qy 2143 ttgacatttagcgaatcaagatacctttaataatagtgtgggttactaaagaagtaaac 2202
Db 1296 caatatcatttcgttcaagaagatgattgtgtgtttatgtatttaagagaagtataa 1355

Qy 2203 gactttctctgtttattttaaacactgttacaggaaaactcgcaaaataaataattact 2262
Db 1356 taaatgatatacttctcttaaaaaaiaaaaaaaaaaaaaaaaaa 1415

Qy 2263 gaaaaaiaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2311
Db 1416 aaaaaaiaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1464

RESULT 7
US-08-545-196B-10
; Sequence 10, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-545-196B-10

Query Match      2.8%; Score 64; DB 3; Length 1582;
Best Local Similarity 65.3%; Pred. No. 3.2e-06;
Matches 94; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 2168 tttaataaatgtgtgggttactaaagaagtaaacgacttctctgtttattttaaca 2227
Db 1436 TTTTCTATCTCTATATGTTTAAAGTATATAATAAAATATTTATTTTAAAAA 1495

Qy 2228 ctgtcagcagaactcgcaaaataaataattactgaaaaaiaaaaaaaaaa 2287
Db 1496 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1555
```


QY 2288 aaaaaaaaaaaaaaaaaaaaaa 2311
|||||
Db 1556 AAAAAAAAAAAAAAAAAAAAAA 1579

RESULT 8

US-08-545-196B-12
; Sequence 12, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:

; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA

; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050

; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-545-196B-12

Query Match 2.8%; Score 64; DB 3; Length 1582;
Best Local Similarity 65.3%; Pred. No. 3.2e-06;
Matches 94; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 2168 tttaataaatatggtgggtactaaagaagtaaacgactctcctgtttattttaaaca 2227

|||||
Db 1436 TTTTCTATCTTATATGTTTAAAGTATATATAAAAAATATTATTTTAAAAA 1495

QY 2228 ctgttcaggaacgcgaattatttactaaataaaataaaataaaataaaataaa 2287

|||||
Db 1496 AA 1555

QY 2288 aaaaaaaaaaaaaaaaaaaaaa 2311
|||||
Db 1556 AAAAAAAAAAAAAAAAAAAAAA 1579

RESULT 9

US-09-817-180-1
; Sequence 1, Application US/09817180
; Patent No. 6340584
; GENERAL INFORMATION:

; APPLICANT: GAN, weiniu et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001183
; CURRENT APPLICATION NUMBER: US/09/817,180
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1

; LENGTH: 2674
; TYPE: DNA

; ORGANISM: Human
US-09-817-180-1

Query Match 2.7%; Score 63.4; DB 4; Length 2674;
Best Local Similarity 70.2%; Pred. No. 6e-06;

Matches 85; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2191 aagaagtaaacgacttctctctgtttattttaaacactgttacaggaactcgcaaaa 2250

|||||
Db 2520 aacataaaaccactgtgccactgaaaaaataaaataaaataaaataaaataaa 2579

QY 2251 ttaataattactgaaaaaataaaataaaataaaataaaataaaataaaataaa 2310

|||||
Db 2580 aaaaaaataaaataaaataaaataaaataaaataaaataaaataaaataaaataaa 2639

QY 2311 a 2311

Db 2640 a 2640

RESULT 10

US-09-027-137-2

; Sequence 2, Application US/09027137
; Patent No. 6013450
; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.

; APPLICANT: Yue, Henry
; TITLE OF INVENTION: CAF1-RELATED PROTEIN

; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.

; CITY: Palo Alto
; STATE: CA

; COUNTRY: USA
; ZIP: 94304

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/027,137
; FILING DATE: Filed Herewith

; CLASSIFICATION:

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0476 US
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 2852 base pairs
; TYPE: nucleic acid

; STRANDEDNESS: single
; TOPOLOGY: linear

; IMMEDIATE SOURCE:

```
; LIBRARY: PROGNOT16
; CLONE: 2229466
; US-09-027-137-2

Query Match          2.7%; Score 63.4; DB 3; Length 2852;
Best Local Similarity 78.4%; Pred. No. 6.2e-06;
Matches 76; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 2215 ttattttaaactgttacaggaaactgcgaataataattactgaaaaaataaa 2274
|| ||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2483 TTCCTTTTAACTGTCGCAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 2542
|| ||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2275 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2311
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2543 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2579

RESULT 11
US-08-821-994-63
; Sequence 63, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 1441
; TYPE: DNA
; ORGANISM: Brassica napus
; US-08-821-994-63

Query Match          2.7%; Score 63.2; DB 4; Length 1441;
Best Local Similarity 71.6%; Pred. No. 4.8e-06;
Matches 83; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2196 agtaaacactctctctgtttattttaaacactgttacaggaaactgcgaataaa 2255
|| ||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1308 attcaagaattgtctgttgctgtattgtattgaagaaataataataaatgata 1367

QY 2256 tattactgaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2311
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1368 tattctcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1423

RESULT 12
US-09-721-822A-10
; Sequence 10, Application US/09721822A
; Patent No. 6306606
; GENERAL INFORMATION:
; APPLICANT: Michael J. Weber
; APPLICANT: Jacqueline Wyatt
; APPLICANT: Lex M. Cowsett
; TITLE OF INVENTION: ANTISENSE MODULATION OF MP-1 EXPRESSION
; FILE REFERENCE: RTS-0142
; CURRENT APPLICATION NUMBER: US/09/721,822A
; CURRENT FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 135
; SEQ ID NO 10
; LENGTH: 1315
; TYPE: DNA
; ORGANISM: Mus musculus
```

```
/ NAME/KEY: CDS
/ LOCATION: 78 to 155
/ LOCATION: 618 to 725
/ LOCATION: 884 to 1018
/ IDENTIFICATION METHOD: by experiment
US-08-909-965C-8

Query Match          2.7%; Score 63; DB 2; Length 1512;
Best Local Similarity 75.7%; Pred. No. 5.9e-06;
Matches 78; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 2209 ttctgttattttaaacactgtacaggaactcgcaaaattaaataattactgaaaaa 2268
      ||| ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 1391 TACCAATATTTCACAAATTTAGAAAAATCCAAAAATCATATGGACCAAAAA 1450

QY 2269 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2311
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1451 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1493

RESULT 14
US-08-242-677-1
; Sequence 1, Application US/08242677
; Patent No. 5677143
; GENERAL INFORMATION:
; APPLICANT: Gaynor, Richard B
; APPLICANT: Wu, Foon W.
; TITLE OF INVENTION: Cellular Nucleic Acid Binding Protein
; TITLE OF INVENTION: and Uses Thereof in regulating Gene Expression and in the
; TITLE OF INVENTION: Treatment of AIDS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/242,677
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mayfield, Denise L.
; REGISTRATION NUMBER: 33,732
; REFERENCE/DOCKET NUMBER: UTSD:401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5173 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4863
US-08-242-677-1

Query Match          2.7%; Score 63; DB 1; Length 5173;
Best Local Similarity 66.7%; Pred. No. 1.1e-05;
Matches 90; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 2173 taaatgtgggttactaaagaactcttctctgtttattttaaacacttgt 2232
```

```
Db 5039 TTAATAAACACTTTTGTGTTAAATTCCTTCCTTTAATAAATATTTTAAGCAATTCT 5098
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2233 acaggaaactcgcaaaattaaataattactgaaaaaataaaataaaataaa 2292
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5099 CCAATAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5158

QY 2293 aaaaaaaaaaaaaa 2307
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5159 AAAAAAAAAAAAAA 5173

RESULT 15
US-09-014-969-14
; Sequence 14, Application US/09014969
; Patent No. 5965397
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/014,969
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-014-969-14

Query Match          2.7%; Score 62.8; DB 2; Length 2447;
Best Local Similarity 56.9%; Pred. No. 8e-06;
Matches 115; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 2110 tcaaaatagaatgacatgtgtgtgatattgacatttagcgaatacaagatacctt 2169
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2140 TGAAGTCTAAATAATGCTGTATAGATATGTATATGTTTCACATATCTGGATCTGTG 2199

QY 2170 taataaatatggtgggttactaaagaactcttctctgtttattttaaacact 2229
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2200 TATTTGATTTTGTACTTTTAAATGTGACAAATAAACCTTTTGGGAGAAAAA 2259

QY 2230 tgtacaggaaactcgcaaaataattactgaaaaaataattactgaaaaaataaa 2289
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 00:11:01 ; Search time 3580.18 Seconds
(without alignments)
8712.261 Million cell updates/sec

Title: US-09-807-470-1
Perfect score: 2311
Sequence: 1 agagagagagagagagag.....aaaaaaaaaaaaaaaaaa 2311

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_gss:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	717.6	31.1	825	10	BI731069	BI731069 603351703
2	699.6	30.3	773	10	BI730813	BI730813 603351811
3	628	27.2	688	9	BB648723	BB648723 BB648723
4	605.2	26.2	651	9	BB641088	BB641088 BB641088
5	594.4	25.7	645	9	BB521515	BB521515 BB521515
6	587	25.4	628	9	BB662449	BB662449 BB662449
7	577.6	25.0	610	9	BE109193	BE109193 UT-R-BS1-
8	534.4	23.1	600	10	BI985062	BI985062 3131-07 M
9	495	21.4	1356	11	BC020283	BC020283 Homo sapi
10	484.2	21.0	1064	9	AL523342	AL523342 AL523342
11	426.6	18.5	1142	10	BM466240	BM466240 AGENCOURT
12	418.2	18.1	622	12	A2958987	A2958987 2M0226C06
13	407.2	17.6	1135	10	BM466239	BM466239 AGENCOURT
14	392.4	17.0	814	9	AU132871	AU132871 AU132871
15	376.2	16.3	822	9	AU130188	AU130188 AU130188
16	373.2	16.1	790	10	BB612552	BB612552 601452065
17	370.6	16.0	944	10	BI655670	BI655670 603283957

18	369.2	16.0	905	9	AU131342	AU131342 AU131342
19	362.6	15.7	836	10	BF969571	BF969571 602271922
20	361.4	15.6	879	10	BI870787	BI870787 603394607
21	361.2	15.6	779	10	BM045930	BM045930 603624955
22	356	15.4	808	10	BI830439	BI830439 603073495
23	354.6	15.3	910	10	BM459490	BM459490 AGENCOURT
24	353.4	15.2	479	9	AI848290	AI848290 UI-M-AH1-
25	352.2	15.2	727	10	BI733351	BI733351 602573813
26	344.8	14.8	695	10	BI733351	BI733351 603353253
27	342.8	14.7	760	10	BI714759	BI714759 602673845
28	340.8	14.7	879	9	AU128453	AU128453 AU128453
29	339.4	14.7	1067	10	BF538214	BF538214 602053686
30	338.4	14.6	636	10	BI8077352	BI8077352 H3014E03
31	337.6	14.6	969	10	BI8077352	BI8077352 H3014E03
32	335.6	14.5	1015	10	BI8077352	BI8077352 H3014E03
33	333	14.4	878	10	BE618241	BE618241 601462496
34	330.2	14.3	651	10	BI8077352	BI8077352 H3014E03
35	330.2	14.3	944	10	BI8077352	BI8077352 H3014E03
36	325.8	14.1	769	10	BI8077352	BI8077352 H3014E03
37	322.6	14.0	859	10	BI8077352	BI8077352 H3014E03
38	321.6	13.9	614	9	AL602739	AL602739 DKE2P686F
39	321.4	13.9	708	10	BE914578	BE914578 601665679
40	320.2	13.9	660	10	BF984410	BF984410 602307936
41	320.2	13.9	736	10	BF141338	BF141338 601789863
42	316.2	13.7	608	10	BI8077352	BI8077352 H3014E03
43	316.2	13.7	693	10	BM486565	BM486565 pgmic.pk0
44	313.4	13.6	835	10	BE887450	BE887450 601509592
45	309.8	13.4	569	9	AI201635	AI201635 qb81d09.x

ALIGNMENTS

RESULT 1

BI731069

LOCUS 603351703F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5359105 5', mRNA linear EST 20-SEP-2001

DEFINITION mRNA sequence.

ACCESSION BI731069.1 GI:15708082

VERSION BI731069

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 825)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM11913 row: f column: 02
High quality sequence stop: 822.

FEATURES

source

1..825

/db_xref="taxon:10090"

/clone="IMAGE:5359105"

/clone_lib="NIH_MGC_94"

/tissue_type="retina"

/lab_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

```
BASE COUNT 173 a 222 c 231 g 199 t
ORIGIN
Query Match 31.1%; Score 717.6; DB 10; Length 825;
Best Local Similarity 93.6%; Pred. No. 4.4e-82;
Matches 771; Conservative 0; Mismatches 49; Indels 4; Gaps 2;

Qy 380 acggcgccggatggggagccggcctggagcgcgcagtcagtcagcagtgccggcgg 439
Db 1 ACGGCGCGGTGATGGGAGACCCCGCCCTGGAGCGCTCAGTGATCAGCGTGGAGCGCG 60

Qy 440 tgaagtcacggatgagtcacgcgcggctggcgcgaagaagcagccaggaaccccg 499
Db 61 TGACTACTGTTGAGTACCGCGCATGGGGCTCCGCAAGAAGAACGCCAGAACCCCGG 120

Qy 500 tgcagcgcgaatctatgctgcagacccacgcggatggtctcctcgtgggcatgt 559
Db 121 TGTTCAGCCACCAATTCATGTCGACAGCCAGCGGATATGGTCTCGTGGCATGT 180

Qy 560 tcttcgtcgtgggaacttattgtccagcgcgaagggctaccctcggggtccagacccttacc 619
Db 181 TCTTCGTGCTGGGGCTTATGTTGAGGGCACCTCCGAGATGCGATCGGCTTCTCACATC 240

Qy 620 tgcagcgcgaatctatgctgcagacccacgcggatggtctcctcgtgggcatgt 679
Db 241 TGCACATGAGTGTCTGTTGCCAGCGGAAGGGCTGCCCTCCGGTCCAGGACCTGTACC 300

Qy 680 attatgggtcaaaatctatgctgcagcagtgcttctacatgctgggtggccatcttc 739
Db 301 ATTATGGGGTCAAAAGATCTGGCCACAGTGTCTTCTACATGCTGGTGGCCATCATCC 360

Qy 740 acgcccaccattcaggagtcagtcagataaagtcagccgagcagtcagtcagccaaag 799
Db 361 ACGCCACCATTCAGGAGTACGTGTAGATAAGCTCAGCAGGAGACTGCAGCTCACCAAG 420

Qy 800 gcaacacaaacaaatgtaagtcagcgcgcagtcagtcagtcagtcagtcagtcagtc 859
Db 421 GCAACAAACAAACTGAATGAGCGCGGCAGCTGAGTGTGTTCTATATAGTTCGGGA 480

Qy 860 tctgggttatcatcttgcctctgagaactgcctctcagaccccaactctattgtgga 919
Db 481 TCTGGGGCATGATCAATCTGGCCTCTGAGAACTGCCTGTGACACCTACACTCTTGTGGA 540

Qy 920 agtctcagcccaacacatgatgacatttcagatgaaatttttctacatctcagctgg 979
Db 541 AGTCTCAGCCCCACAAATGATGACATTCAGATGAATTTTCTTACATTTACAGTTGG 600

Qy 980 ctactggttccatagtttcccgagctctactctccagaaagtcaggaacaaagatacc 1039
Db 601 CTTACTGTTTTCATAGCTTTCCGGAGCTTACTTCCAGAAAGTCAGGAAGCAAGATATCC 660

Qy 1040 cgggtcaactcatctacattggcctccacctctccacattggagggccctatcttgt 1099
Db 661 CGGTCAGCTCATCTACATGTCCTCCACCTCTTCCCATTTGAGGGGCCATCTCTTGT 720

Qy 1100 ac-ttgaaccacctggcgtcgtctctgatgctgca---ctatgctgtcgagctctc 1155
Db 721 ACTTTGAACCACTTGGGCTGCTGCTGCTGATGCTGCAACTACGCCGTAGAGGCTCCTC 780

Qy 1156 tccagcgtgtcagcgccttcttaactttgggatgagcgggtacca 1199
Db 781 TCCAGCGTGTGACGCTTGCTTTATTTTGGTGATGAGCGGTACCA 824
```

```
RESULT 2
BI730813 603351811f1 NIH_MGC_94 Mus musculus cdna clone IMAGE:5359106 5',
LOCUS mRNA sequence.
DEFINITION BI730813
ACCESSION BI730813
VERSION BI730813.1 GI:15707826
KEYWORDS EST.
```

```
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 773)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1913 row: f column: 03
High quality sequence stop: 773.
FEATURES
Location/Qualifiers
source 1..773
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5359106"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
Average insert size 3.3 Kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 164 a 210 c 217 g 182 t
ORIGIN
Query Match 30.3%; Score 699.6; DB 10; Length 773;
Best Local Similarity 95.4%; Pred. No. 8.5e-80;
Matches 731; Conservative 0; Mismatches 34; Indels 1; Gaps 1;

Qy 380 acggcgccggatggggagccggcctggagcgcgcagtcagtcagcagtgccggcgg 439
Db 1 ACGGCGCGGTGATGGGAGACCCCGCCCTGGAGCGCTCAGTGATCAGCGTGGAGCGCG 60

Qy 440 tgaagtcacggatgagtcacgcgcggctgcgaagaagcagccaggaaccccg 499
Db 61 TGACTACTGTTGAGTACCGCGCATGGGGCTCCGCAAGAAGAACGCCAGAACCCCGG 120

Qy 500 tgcagcgcgaatctatgctgcagacccacgcggatggtctcctcgtgggcatgt 559
Db 121 TGTTCAGCCACCAATTCATGTCGACAGCCAGCGGATATGGTCTCGTGGCATGT 180

Qy 560 tcttcgtcgtgggaacttattgtccagcgcgaagggctaccctcggggtccagacccttacc 619
Db 181 TCTTCGTGCTGGGGCTTATGTTGAGGGCACCTCCGAGATGCGATCGGCTTCTCACATC 240

Qy 620 tgcagcgcgaatctatgctgcagacccacgcggatggtctcctcgtgggcatgt 679
Db 241 TGCACATGAGTGTCTGTTGCCAGCGGAAGGGCTGCCCTCCGGTCCAGGACCTGTACC 300

Qy 680 attatgggtcaaaatctatgctgcagcagtgcttctacatgctgggtggccatcttc 739
Db 301 ATTATGGGGTCAAAAGATCTGGCCACAGTGTCTTCTACATGCTGGTGGCCATCATCC 360

Qy 740 acgcccaccattcaggagtcagtcagataaagtcagccgagcagtcagtcagccaaag 799
Db 361 ACGCCACCATTCAGGAGTACGTGTAGATAAGCTCAGCAGGAGACTGCAGCTCACCAAG 420

Qy 800 gcaacacaaacaaatgtaagtcagcgcgcagtcagtcagtcagtcagtcagtcagtc 859
Db 421 GCAACAAACAAACTGAATGAGCGCGGCAGCTGAGTGTGTTCTATATAGTTCGGGA 480

Qy 860 tctgggttatcatcttgcctctgagaactgcctctcagaccccaactctattgtgga 919
Db 481 TCTGGGGCATGATCAATCTGGCCTCTGAGAACTGCCTGTGACACCTACACTCTTGTGGA 540

Qy 920 agtctcagcccaacacatgatgacatttcagatgaaatttttctacatctcagctgg 979
Db 541 AGTCTCAGCCCCACAAATGATGACATTCAGATGAATTTTCTTACATTTACAGTTGG 600

Qy 980 ctactggttccatagtttcccgagctctactctccagaaagtcaggaacaaagatacc 1039
Db 601 CTTACTGTTTTCATAGCTTTCCGGAGCTTACTTCCAGAAAGTCAGGAAGCAAGATATCC 660

Qy 1040 cgggtcaactcatctacattggcctccacctctccacattggagggccctatcttgt 1099
Db 661 CGGTCAGCTCATCTACATGTCCTCCACCTCTTCCCATTTGAGGGGCCATCTCTTGT 720

Qy 1100 ac-ttgaaccacctggcgtcgtctctgatgctgca---ctatgctgtcgagctctc 1155
Db 721 ACTTTGAACCACTTGGGCTGCTGCTGCTGATGCTGCAACTACGCCGTAGAGGCTCCTC 780

Qy 1156 tccagcgtgtcagcgccttcttaactttgggatgagcgggtacca 1199
Db 781 TCCAGCGTGTGACGCTTGCTTTATTTTGGTGATGAGCGGTACCA 824
```

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES
source

Location/Qualifiers

1..688
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="C130035P06"
/clone_lib="RIKEN full-length enriched, 16 days embryo head"
/sex="mixed"
/tissue_type="head"
/dev_stage="16 days embryo"
/lab_host="DH10B"
/note="Site.1: Salt; Site.2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of genome exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCACAGACTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATTCGAGTGAATTAATTAATCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda F1C 1"

BASE COUNT 150 a 182 c 202 g 153 t 1 others
ORIGIN

Query Match 27.2%; Score 628; DB 9; Length 688;
Best Local Similarity 95.4%; Pred. No. le-70;
Matches 657; Conservative 0; Mismatches 31; Indels 1; Gaps 1;

QY 359 ggctgcgaccttggcggcagacggcggggatggggagcccgctgggagcgcgca 418
Db 1 GGTGCGACCCCTGGGGCTCAGACGGCGGTGATGGGAGCCCGCGCTCGGAGCGGCTCA 60

QY 419 gtgatcagcgggtgcggcggtgagTaccggtgagTaccgcgcatcgggctccgaaga 478
Db 61 GTGATCAGCGGTGAGCGCGGTGACTACTGTTGATGAGTACCGCGGCATGGGGCTCGGAAGA 120

QY 479 agaacgccaggaaacccccgggtgctgagccacgaattcatgtgtgcagaaacacgcggata 538
Db 121 AGAAGCCAGGAACCCCGCGTGTGAGCCACGAATTCATGTGTGACAAACACACCGCGATA 180

QY 539 tggctctctcgtggcgcatgttcttgcgtgggaactattgttcaggagcgacggcgag 598
Db 181 TGGTCCTCCGTCGCGGGCATGTTCTTCGTCTGGGGCTTATGTTCGAGGGCAC-CCGAGA 239

QY 599 tgcgcagctgttccctcaccctgcagcagtgagctgtgtcccgagcggaaggctaccct 658
Db 240 TGTCGATCGCTTCCTCACTCTGCACATGGAGTCGTGTCCACGGGAGGGCTCGCCT 299

QY 659 cgggggtccaggaccctttaccattatggggtcaaatgctggtccacagtgttcttaca 718
Db 300 CCGGTCACAGGACCCGTACCATATGGGTCAAGATCTGGCCACAGTGTCTTCTTACA 359

QY 719 tgcgtggcccatcatcattcaacgacacattcaggagtagctgtctagataaagctcagcc 778
Db 360 TCGTGTGGCCATCATCATCCAGCCACCATTCAGGAGTACGTGCTAGATAAGCTCAGCA 419

QY 779 ggagactcgagctcaccagggcaaaacaaaattgaatgagcgcgagcagctagtg 838
Db 420 GGAGACTGCAGCTCACAAAGGCCAAACAAACAACTGAATGAGGCCGGCAGCTGAGTG 479

QY 839 tgttctacatagtgctcgtgtatctgggtgatgatcatcttgcctctgagaactgctgt 898

[illegible]

SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 610)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 9704477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoaresblue.weeg.uiowa.edu
 The sequence contained an oligo-dr track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dr track served to verify it as a clone from the
 normalized embryo at 13 dpc library cDNA library Preparation: M.B.
 Soares Lab Clone Distribution: clones will be available through
 Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES
 source Location/Qualifiers
 1. .610
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BS1-azd-a-04-0-UI"
 /clone_lib="UI-R-BS1"
 /dev_stage="embryonic 13 dpc"
 /lab_host="DHI0B (Life Technologies)"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BS1
 library is derived from 13 dpc whole embryo tissue. For a
 detailed description of the library from which this clone
 was derived, please visit our web site at
 ratest.eng.uiowa.edu.
 TAG_LIB="UI-R-BS1"
 TAG_TISSUE="embryo at 13 dpc"
 TAG_SEQ="AATCC"
 BASE COUNT 206 a 118 c 95 g 190 t 1 others
 ORIGIN

Query Match 25.0%; Score 577.6; DB 9; Length 610;
 Best Local Similarity 98.3%; Pred. No. 2.6e-64;
 Matches 594; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 1528 ccaagaagaagaagctccttagcagtgcaagcgaattgattcttaccctccaag 1587
 |||||
 DB 610 CCAAGAAGAAGAAGAGAGCTCTTAGCAGTTGCAAGGATGATTCTTACCTCCAAG 551
 |||||

QY 1588 gaatccactttcttatgtggtgtctctgtctagagatttctgtcttcagaacggg 1647
 |||||
 DB 550 GAATCCACTTCTTCTTATGTGTGTCTCTGTGCTAGAGATTTCTGTCTTCAGAACGGG 491
 |||||

QY 1648 tcgtgcttttgaattgtcattgtattgtcattgtttttaaagtttttcagaagt 1707
 |||||
 DB 490 TCGTGCCTTTTGAATATGCTTAATGTATGTCTAATGTGTTTAAAGTTTTCAGACGCT 431
 |||||

QY 1708 atgagtggggatgggggttaagactaaaccactcagcctctaatacagtcagaatagt 1767
 |||||
 DB 430 ATGAGTGGGGAGGGGGTGAAGACTAAACCACCTCAGCCTTAAACTGTGCAAAAGT 371
 |||||

QY 1768 taacggaccacaatcttatttagtttagtttttaccctcaacgatttccaaacgtttgt 1827
 |||||
 DB 370 TAACGGACCAACATCTTATTAGTTAGTTCTTACCTCAACGATTTTCCAAACGTTTGT 311
 |||||

QY 1828 ggtgatgactcagaaatgtgtacataaataatagtttctcgttcacaaatgtctttatc 1887
 |||||
 DB 310 GGTGATGACTCGAGAAATGTGTACATAAATAATAGTTCCTGCTTCCAATGTCCTTATC 251
 |||||

QY 1888 gaattaacaagtctgtcagcaaaagtgtgttttttctcaatgttctcctgcaggataaag 1947
 |||||
 DB 250 GAATTAACAAGTCTGCTAGCAAAAGTGTGTTTCTCAATGTCTCTCTGCAGATAAAG 191
 |||||

QY 1948 tggaaaatctgataaaggttaaaactcaaatcagttattatgtaaccgttgggatttttta 2007
 |||||
 DB 190 TGGAAAATCTGATAAAGGTAAACCTCAATCAGTATATATGTAACCGTTCGGGATTTTTTA 131
 |||||

QY 2008 aagtttttaaatattacaaatgaaagcattgtcacaaccac-caaaaaatgtgttaat 2066
 |||||
 DB 130 AAGTGTTTTAAATTTACAAATGGAAGCATTTGTCAAAACCCACNCAAAATATGTGTTAAT 71
 |||||

QY 2067 tttatgagtagtaattgttagtgcttacccccccattaaagcatcaaaatatgaatagat 2126
 |||||
 DB 70 TTTATGAGTAGTAATTTGTAGTGTCTTACGCCCCCAATTAAGCATCAAAATTTGAAAAAA 11
 |||||

QY 2127 gaca 2130
 |||
 DB 10 AAAA 7
 |||

RESULT 8
 BI985062 BI985062 600 bp mRNA linear EST 20-DEC-2001
 LOCUS 3131-07 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
 DEFINITION mRNA sequence.
 ACCESSION BI985062
 VERSION BI985062.1 GI:17955990
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 600)
 AUTHORS Mu X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
 White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
 TITLE Gene expression in the developing mouse retina by EST sequencing
 and microarray analysis
 JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)
 COMMENT Contact: Klein WH
 Department of Biochemistry and Molecular Biology
 University of Texas M.D. Anderson Cancer Center
 Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
 Tel: 713 792 3646
 Fax: 713 790 0329.

FEATURES
 Location/Qualifiers
 source 1. .600
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
 /tissue_type="neural retina"
 /dev_stage="embryonic day 14.5 post-fertilization"
 BASE COUNT 131 a 159 c 168 g 140 t 2 others
 ORIGIN

Query Match 23.1%; Score 534.4; DB 10; Length 600;
 Best Local Similarity 95.3%; Pred. No. 7.8e-59;
 Matches 572; Conservative 0; Mismatches 26; Indels 2; Gaps 2;

QY 409 gagcgcgcagtgatcagcgggtggcgccggtgagtcaccggtgagtcaccgcgcagtgagg 468
 |||||
 DB 1 GAGCGGCTCAGTGATCAGCGGTGGAGCGCGGTGACTACTGTGTTGAGTACCGCGCATGGG 60
 |||||

QY 469 ctcccgcaagaagacgcaggaaccccccggtgctgagcac-gaattcatggtgcagaa 527
 |||||
 DB 61 CTCGCCAAGAGAAGCCAGGACCCCGCGGTGTGAGCCACNGAATTCATGTGTGCAGAA 120
 |||||

```
QY 528 ccacgc-ggataggctctcgtcgtaggcagctgtcttcgtgctgggacttatgttcgag 586
Db 121 CCACGCGGATATGGTCTCTCGCTGGGCATGTTCTTCTGCTGGGCTTATGTTTCGAGG 180
QY 587 gcacgcccagagatgctgatcgtgttctcctcacccctcagcatgagtcgtttgtcccccagcg 646
Db 181 GCACCTCCGAGATGTCGATCGGTTCTCTCACCTCTGCAGCATGGAGTCGTTGTGCCAGCGG 240
QY 647 aagggtacccctcggggtccagagaccctttaccattatggggtcaaatgctcgtccacag 706
Db 241 AAGGGTGCCTCCGGGTCGAGGACCTGTACCATTTATGGGGTCAAGAGTCTGGCCACAG 300
QY 707 tgtttctacatgctggtggccatcatcttcacccacccattcaggagtcgtgctag 766
Db 301 TGTCTCTACATGCTGGTGGCCATCATATCCAGCCACCATTCAGGAGTAGCTGCTAG 360
QY 767 ataagctcagcggagactgcagctcaccaaaaggcgaacaaacaaattgaatgagccg 826
Db 361 ATAAGCTCAGCAGGACTGCAGCTCACCAAGGCAACAAACAACTGAATGAGGCGG 420
QY 827 ggcagctgagtgctctacatagtgctggtatctggtatctggggtatgatcttctgacctcg 886
Db 421 GGCAGCTGAGTGTCTATATAGTGTCCGGGATCTGGGGCATGATATTTCTGGCCCTCG 480
QY 887 aqaactgcctgcagaccacactctattgtggaagtctcagcccccacacatgatgacat 946
Db 481 AGAAGCTGCTGTGAGACCTTACACTCTTGTGGAAGTCTCAGCCGCCAACACATGATGACAT 540
QY 947 ttcagatgaattttctacatctcagttcagttggttactggtttcattagtttccccgagc 1006
Db 541 TTCAGATGAATTTTCTACATTTTCAGATTGCGTGGCTTACTGGTTTCATGAGCTTCCCGGAGC 600

RESULT 9
BC020283 1356 bp mRNA linear HTC 20-DEC-2001
LOCUS Homo sapiens, clone IMAGE:4519858, mRNA.
DEFINITION BC020283
ACCESSION BC020283
VERSION BC020283.1 GI:17946782
KEYWORDS HTC.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1356)
Strausberg, R.
Direct Submission
Submitted (20-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: DCTD/dTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalona@bcm.tmc.edu.
Villalona, D.K., Luna, R.A., Hale, S.M., Rulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gunaratne, P., Yoon, V., Kowis, C., Martin, R.,
Lawrence, S., Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 29 Row: h Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
```

This clone has the following problem: no 5' EST match.

```
FEATURES
    source          Location/Qualifiers
     1..1356
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4519858"
        /tissue_type="Prostate, adenocarcinoma."
        /clone_lib="NIH_MGC_91"
        /lab_host="DH10B"
        /note="Vector: pCMV-SPORT6"
BASE COUNT      401 a   279 c   268 g   408 t
ORIGIN
Query Match      21.4%;   Score 495;   DB 11;   Length 1356;
Best Local Similarity 65.0%;   Pred. No. 6.e-54;
Matches 825;   Conservative 0;   Mismatches 410;   Indels 34;   Gaps 5;
QY 436 ccggtgagtagcggtagtaccgcgccatgggggtccgcaagaagacccaggaacccc 495
Db 49 CCTGCAGCGCGGGCGGGCCACCATGGCGATTTCGAAGAAAAGCACCAGAGCCGCC 108
QY 496 ccggtgtagcagcaaatcattcagtgacagaccacgcgcatatggtctcctcgtggtggc 555
Db 109 CCAGTGTGTAGCCAGCAATTCGTCTCGAGAAATCACGGGACATCGTCTCTGTGTGGCG 168
QY 556 atgtcttcgtgctgggacttatgttcgagggccagccgagagatgctcagatcgtgttcctc 615
Db 169 ATGGTCTTCTGCTGGGGCTCATGTTTGAGATAAAGGCAAAAGCTTCTATCATTTTGTGT 228
QY 616 accctgcagcatggagtcgttggctccag ---cggaaagggtaccctcgggggtccaggacc 672
Db 229 ACTCTTCAGTACAATGTACCCCTCCAGCAACAGAAAGCAAGCTPACTGAATCAGTGCCTC 288
QY 673 ctttaaccattatgggggtcaaatgctggccacagtgttcttctacatcgtgtggccatc 732
Db 289 CTTTATTACTATGGCATCAAAAGATTTGGCTACTGTTTCTTCTACATGCTAGTGGCGATA 348
QY 733 atcattcacgcaccattcagggagtcagctgctagataagctcagccgagagactgcagctc 792
Db 349 ATTATTATGCGCGTAATTCGAAGAGTATATGTTGGATAAAATTAACAGGCGAATGCACCTTC 408
QY 793 accaaaggcaacaaacaaatgaatgagccggcgagctgagtggtttctacatagtg 852
Db 409 TCCAAAACAAAACACAGCAAGCTGAATGAATCTGTGCTAGCTTAGTGCGTCTTACCTTTT 468
QY 853 tctggtatcctgggtgatgatcatctcctcctgagaaactgctgctcagacccccactcta 912
Db 469 GTCGTGTTTGGGGCACATTTCATCTCATCTCTGAAAGCTACATCTCAGACCCCACTATC 528
QY 913 ttgtggaagtctcagcccccacacatgatgacatttcagatgaaatttttctacatctca 972
Db 529 TTATGGAGGGCTTATPCCCATAACTGATGATTCATTAATGAAGTTTTTCTACATATCA 588
QY 973 cagttggcttactggtttcactagtttccggagctctacttccagaaagtcaggaaacaa 1032
Db 589 CAGCTGGCTTACTGCTTCATGCTTTTCTGTAACCTTACTTCCAGAAAACCAAAAAGAA 648
QY 1033 gatattccgggttcaactcatctacattggcctccaccctcttccacattggaggggacctat 1092
Db 649 GATATCTCTGCTGAGCTTGTCTACATTGGTCTTTTACCTCTTCCACATTCGTGGAGCTTAC 708
QY 1093 cctctgacttgaacacacactgggcctgctgcttctcagatgctcactatgctcagacctc 1152
Db 709 CTTTGAACCTTGAATCATCTAGGACTTGTCTCTGGTGTCTACATTTTGTGTAATTT 768
QY 1153 ctcctcagcgtgtagcctgctttactttggggatgagcgtaccagaaagggtgtctc 1212
Db 769 CTTTTCACATTTCCGCGCTGTTTATTTTACATGAAGATATACAGAAAGGATTTTCT 828
QY 1213 ttgtggcctatcgtgttttatctatccgggagactcgtgcacactgattgtctcagtggttaca 1272
Db 829 CTGTGGCGAGTCTCTTTTGTGTTGGGAGAGACTTCTGACTTTAATTTCTTCTCAGTACTGACT 888
```

```
QY 1273 gtaggcttcacttgcccgagac---aaatcggaatgaaatgctctctctgtaatgtc 1329
Db 889 GTGGTTTGGCCTTCGAAGAGCAGAAATCAGAAGCTGGATTTCAGTACTGGAACATTC 948
QY 1330 aatgtgttgccagcagtaaatcgctgtctgtctcgcagttgcagtatccagggtgtacata 1389
Db 949 AATGTGTTAGCTGTTAGAAATCGCTGTCTGGCATCCATTGCGTTACTCAGGCATTTATG 1008
QY 1390 acatggacottgacgaccgtctggtctcagagatggttgaagaatcggaatcttcattgtc 1449
Db 1009 ATGTGGAAGTTCAATTAATTTTCAGCTTCGAAGGTGGAGGGAACATTCTGCTTTTCAGCA 1068
QY 1450 tftgggaggaag-----agacggtccaggtcgagaaagccacagaa 1491
Db 1069 CCAGCTGTGAAGAGAAACCAACAGTAACATAAGGCAGATCTTCTAATAAAGAAACAGAA 1128
QY 1492 aatggagtga-----gaatccaaatagaaatgattctccacaaagaagaagag 1542
Db 1129 AATGGTGTGAATGGAACATTAACCTCAATGTAGCAGACTCTCCCGGAATAAAAAAGAG 1188
QY 1543 aaagc-tcattagcaattgcaagcgaattgattcttacctccaaggaatccacttcttc 1601
Db 1189 AAATCTTCATAATGAATTAATAACTAATTAATTAATGTCGCCAAAGAAATCTGCTTTCTA 1248
QY 1602 ttatgtggtctctgctgtagagatttctgttcttcagaacggtcgctgctttttgaa 1661
Db 1249 CTATATCTTTCAGCANTAGAGATTTTCTGTTCTTGAAATACAGTCTGTGCTCTTGAT 1308
QY 1662 tattgctaa 1670
Db 1309 TTTTAAAAA 1317

RESULT 10
AL523342
LOCUS
DEFINITION AL523342 LFI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC001YH18 5
prime, mRNA sequence.
ACCESSION AL523342
VERSION AL523342.1 GI:12786835
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1064)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..1064
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC001YH18"
/clone_lib="LFI_NFL003_NBC3"
/sex="male"
/tissue.type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
```

```
BASE COUNT 270 a 236 c 220 g 336 t 2 others
ORIGIN
Query Match 21.0%; Score 484.2; DB 9; Length 1064;
Best Local Similarity 68.7%; Pred. No. 1.6e-52;
Matches 695; Conservative 1; Mismatches 309; Indels 6; Gaps 2;
QY 426 gcggtggcgccggtgagtaccggtgagtcacgcgagatgggctccgcaagaagaacgc 485
Db 1 GGGGCGSAGACCTTCAGCGGGGGGGGGCCACCATGGCGATTCGCAAGAAAGAC 60
QY 486 cagaaaccccccggtgctgagccacgaattcatggtgcagaacacgcggtataggtctc 545
Db 61 CAAGAGCCCCCAGTGTGAGCCACGAATTCGTCGAGAATCACGGGACATCGTCTC 120
QY 546 ctgcgtgggcatgtcttcgtggtggaacttatgttcgagggcgccgagatgtcagat 605
Db 121 CTGTGTGGCATGTGCTTCTCTGCTGGGCTCATGTTTGAGATAACGGCAAAAGCTTCTAT 180
QY 606 cgtgttctcacccctgcagcatggagtcgttgtcccaag---cgaaagggtaccctcggg 662
Db 181 CATTTTGTGTTACTCTTCAGTACAATGTACCCCTCCCAACAACAGAAGAACAGTACTGA 240
QY 663 gtccaggaacctttaccattatgggtcaagatctggccacagtggtctctctcatgct 722
Db 241 ATCAGTGTCCCTTTTATTACTATGGCATCAAGATTTGGCTACTGTTTCTTCTACATGCT 300
QY 723 ggtggccatcatctacgcgccattccagagtagctgtagatgaagtcagccgag 782
Db 301 AGTGGCGATAAATTATCATGCGTAATTCAAGAGTATATGTTGGATAAAATTAACAGCGC 360
QY 783 actgcagctcacaaaggcaaaacaaataatgaagccgagccgagctgagtggttt 842
Db 361 AATGCACCTCTCCAAAACAAACACAGCAAGTTTAATGAATCTGTCTAGCTAGTGCCTT 420
QY 843 ctacatagtgctgtgattctggtggtatgatacttctgacctctgagaactgcctgtcaga 902
Db 421 CTACCTTTTGGCCTGTGTTGGGGCACATTCATTTCTCATCTCTGAAAACATACATCTCAGA 480
QY 903 ccccaactctattgtggaagtcctcagcccaacacatgatgacatttcagatgaaatttt 962
Db 481 CCCAACTATCTTATGAGGGGCTTATCCCATCAACCTGATGACATTTCAAAATGAAGTTTTT 540
QY 963 ctacatctcacagttgcttactggtttcatagtttccggagctctacttcccaagaagt 1022
Db 541 CTACATATCATCAGCTGGCTTACTGGCTTCTACTGCTTTTCTCTGCTGCTACATATTATT 600
QY 1023 cagaaacaagatatcccggtggtcaactcatctacattggcctccacctcttccacattgg 1082
Db 601 CAAAAAAGAGATATTCTCTGCTAGCTGTCTACATTGGTCTTTTACCCTTCCCATTTGC 660
QY 1083 aggggcctatctctgttacttgaaccacctgggcctgctgcttctgtatgctgcactatgc 1142
Db 661 TGGAGCTTACCTTTTGAACCTTGAATCATCTAGGACTTGTCTCTGCTGCTACATATTATT 720
QY 1143 tgtcagctctctccagcggtgcagcctcttacttacttgggagatgagcgagtaacaa 1202
Db 721 TGTGTAATTTCTTTTCCACATTTCCCGCTGTTTTATTATTAGCAATGAAGAAATATCAGAA 780
QY 1203 aggggtgtcttctgtgacctatcggttttatatccgggagacctcgtagactgattgtctc 1262
Db 781 AGGATTTTCTGTGGGCGAGTCTTTTGTGTTTGGGAAGACTTCTGACTTTTAATTTCTTC 840
QY 1263 agtgggttacagtaggggttacttgcgcgggac---aaatcggaatggaatgctctctc 1319
Db 841 AGTACTGACTGTTGGTTTGGCTTTCGAAGAGCAGAAAAATCAGAAGCTGGATTTTCAGTAC 900
QY 1320 tggtaatgtaaatgtgttggcagctaaaatcgctgttctgctcagattcagtatcca 1379
Db 901 TGGAAACTTCAATGTGTAGTGTAGAACTGCTGTTCTGTCGATCCATTTGCGTTACTCA 960
QY 1380 ggtgtacataacatggaccttgacgacctctctggtcttcagagatggttaga 1430
```

```

Db 961 GCATTATGATGGAGGTCATTAAATTTTCAGCTTCGAAGGTGGAGGA 1011
|||||
Db 325 ATTATTCATGCCCTAATTCAGAGGTATATCTTGGATAAAATTAACAGCGCAATCACTTC 384
|||||
QY 793 accaaaggcaacaaacaaattgaatgagccggcagctgagtggttctacatagtg 852
|||||
Db 385 TCCAAACAAACAAACACAGCAGGTTTAATGAATCTGGTCAGCTTAGTGGTGTCTACCTTTT 444
|||||
QY 853 tctgggtatctgggggtatgatcattctggcctctgagaactgcctgtcagaccccaactca 912
|||||
Db 445 GCCTGTGTTGGGGCACATTCATCTCACTCTGAAAACTACATCTCAGACCCCAACTATC 504
|||||
QY 913 ttgtggaagtctcagcccccacacatgatgacatttcagatgaatttttttcatatctca 972
|||||
Db 505 TTATGAGGGCTTATCCCATAACTGATGACATTTCAATGAAGTTTTTCTTACATATCA 564
|||||
QY 973 cagttggctactggttttcaatagtttcccgagagctctactccagaaagtcaggaacaa 1032
|||||
Db 565 CAGCTGGCTTACTGGGCTTCATGCTTTCTGAACTCTACTTCCAGAAAAACCAAAAAAGAA 624
|||||
QY 1033 gatatcccggtcaactcatctacattggcctccacctctccacattggaggggcctat 1092
|||||
Db 625 GATATTCCTCGTCAGCTGTCTACATTTGCTTTTACCCTCTCCACATTTGCTGGAGCTTAC 684
|||||
QY 1093 ctcttacttggaacacacctggcctgctgctctctgtgactgtgactatgtgtcgagctc 1152
|||||
Db 685 CTTTTGAACCTGAATCATCTAGGACTTGTCTTCTGGTGCTACATTTATTTGTGAATTT 744
|||||
QY 1153 ctctccagcgtgctgagcctgcttactttggggatgagcgtaccagaagggtgtct 1212
|||||
Db 745 CTTTTCACATTTCCCGCCTGTTTATTTAGCAATGAAAGTATCATGAAAGGATTTTCT 804
|||||
QY 1213 ttgtgacctatcgtgtttatctcgggagagactcgtgacactgatt 1257
|||||
Db 805 CTGTGGGCACTTCTTTTGTGTTGGGGAAGACTCTCTGACTTTAAAT 849
|||||

RESULT 12
AZ958987
LOCUS
DEFINITION
2M0226C06R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
clone UUGC2M0226C06 R, DNA sequence.
ACCESSION
AZ958987
VERSION
AZ958987.1 GI:13830214
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 622)
AUTHORS
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D.,Weiss,R.
TITLE
Mouse whole genome scaffolding with paired end reads from 10kb
Plasmid inserts
JOURNAL
Unpublished (2000)
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Em. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0226 row: C column: 06
Seq primer: CACACAGGAACAGCATGACC
Class: plasmid ends
High quality sequence stop: 622.
Location/Qualifiers
1..622
/organism="Mus musculus"
/strain="C57BL/6J"

FEATURES
source
```

RESULT	13
BM466239	
LOCUS	1135 bp mRNA linear EST 05-FEB-2002
DEFINITION	ACENGCURT_6457038 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5576991
	5', mRNA sequence.
ACCESSION	BM466239
VERSION	BM466239.1 GI:18515281
KEYWORDS	EST.

```

Db 506 TTATGGAGGCTATCCCATCAACCTGATGACATTTCAAAATGAAGTTTTTCTACATATCA 565
QY 973 cagttgctactgtttccatagtttccggagctctacttccagaadtccaggaacaa 1032
Db 566 CAGCTGGCTACTGGCTTTTCATGCTTTTCTGAACTCTACTCCAGAAAACCAAAAAGAA 625
QY 1033 gataccgggtgcaactcatctacattggcctccactcttccacattggagggcctat 1092
Db 626 GATATTCCTCGTCAGCTTGCTACATTGGCTTTACCTCTTCCACATTGCTGGAGCTTAC 685
QY 1093 ctcttgtaactgaacacactggcctgctgctctgtatgctgactatgctcgagctc 1152
Db 686 CTTTGAACCTGAATCATCTAGGACTTGTCTTCTGGTCTACATTAATTGTGTAATT 745
QY 1153 ctctccagcgtgcagcgtcttactttgggat-gagcgtaccagaaagg--tt 1208
Db 746 CTTTCCACATTTTCGGCTGTTTATTTTACAAATGGAAGATACAGAAAGGGATTTT 805
QY 1209 gtctttggcctatcgtttatatccgggagactcgtgacactgattgtctcagtggt 1268
Db 806 CTTCTGGGCCAGNCTTTTGTGTTTTTGGGAAACCTCTGACTTTTAATCCTTTCAGAACT 865
QY 1269 t 1269
Db 866 t 866

RESULT 14
AUI32871 AUI32871 814 bp mRNA linear EST 24-OCT-2000
LOCUS AUI32871 NT2RP4 Homo sapiens cDNA clone NT2RP4000757 5', mRNA
DEFINITION sequence.
ACCESSION AUI32871
VERSION AUI32871
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 814)
Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J.,
Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuho,Y. and
Isogai,T.
HRI human cDNA project (Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y.,
Saito,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano
S., Masuho,Y., Isogai,T.)
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute
Location/Qualifiers
1. 814
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP4000757"
/clone_lib="NT2RP4"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18FL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 206 a 195 c 161 g 249 t
ORIGIN
```

```

Query Match 17.0%; Score 392.4; DB 9; Length 814;
Best Local Similarity 70.7%; Pred. No. 7.5e-41;
Matches 550; Conservative 0; Mismatches 223; Indels 5; Gaps 2;

QY 436 ccggtgagtagtaccggtgagtagtaccggtggtccgcgaagaagaagccaggaacccc 495
Db 18 CTTGCAAGGGCGGGCGGGCCACCATGGCGATTTCGAAGAAACGACCAAGAGCCCC 77
QY 496 ccggtgtagccacacgaattcatgtgcagaaaccacgcgatatggtctcctcgctgggc 555
Db 78 CCAAGTCTGAGCCACAGAAATTCGCTGCAGAAATCACGGGACATCGTCTCTGTGTGGCG 137
QY 556 atgtcttgctgctggagacttatgttcgagggcacggccgagatgtcgatgtcttcctc 615
Db 138 ATGGTCTTCTGCTGGGGCTCATGTTTGAGATAACGCAAAAGCTTCTATCATTTTGT 197
QY 616 accctgcagcatgagtcggtgtccag---cggaaaggctacccctgggggtccagacc 672
Db 198 ACTCTTCAAGTACAAATGTCACCTCCCAAGCAACAGAAAGTACTGAAATCAGTGTC 257
QY 673 ctttacattatgggtccaaagatctgtgccacagtgcttctctacatgctggtggccatc 732
Db 258 CTTTATTACTATGGCATCAAGATTTGGCTACTGTTTTCTTCTACATGCTAGTGGCGATA 317
QY 733 atcattcagccaccattcaggagtagtgcgtagatagaagctcagccggagactcagctc 792
Db 318 ATTATTTCATGCCGTAATTCAGAGTATATGTTGGATAAAATTAACAGCGCAATGCAC 377
QY 793 accaaaggcaaaacaaatgaatgagccggcgagctgagctggttctacatagtg 852
Db 378 TCCAAAACAAACACACAGCAAGTTTAAATGAACTGTGGTCAGCTTAGTGGCTTCTACCT 437
QY 853 tcgtgatctggggtatgattccttgcctcgtgagacgctcgtcagaccccaactcta 912
Db 438 GCCTGTGTTTGGGGCACATTCATCTCTGAAAACTACATCTCAGACCCCAACTATC 497
QY 913 ttggaagtctcagccccacaacatgatgacatttcagatgaatatttttctacatctca 972
Db 498 TTATGGAGGGCTTATCCCATAACTGATGACATTTCAAATGAAAGTTTTTCTACATATCA 557
QY 973 cagttgcttactggtttcctagtttcccgagctctacttccagaagtcagagaacaa 1032
Db 558 CAGCTGGCTTACTTGGCTTTCATGCTTTTCCGAACTCTACTTCCAGAAAACCAAAAAGAA 617
QY 1033 gataccgggtgcaactcatctacattggcctccactcttccacattggaggggcctat 1092
Db 618 GATATTCCTCGTCAGCTTGCTACATTTGGTCTTTACCTCTTCCACATTTGCTGGAGCTTAC 677
QY 1093 ctcttgtaactgaaccacactggcctgctgctctctgtatgctgcactatgctgcagctc 1152
Db 678 CTTTGAACCTGAATCATCTAGGACTTGGTCTTCTGGTGTGCTACATTTAT--TTGTGAATTT 735
QY 1153 ctctccagcgtgtagcgtcttactttggggatgagcgggtaccagaaagggtgtg 1210
Db 736 CTTTTCACATTTCCCGCCTGGTTTATTTAGCANTGAAAAGTATCAGAAANGGATTT 793

RESULT 15
AUI30188 AUI30188 822 bp mRNA linear EST 24-OCT-2000
LOCUS AUI30188 NT2RP3 Homo sapiens cDNA clone NT2RP3000388 5', mRNA
DEFINITION sequence.
ACCESSION AUI30188
VERSION AUI30188
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 822)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 14:51:51 ; Search time 74.4 seconds
(without alignments)
541.933 Million cell updates/sec

Title: US-09-807-470-2

Perfect score: 1892

Sequence: 1 MGLRKNARNPVLSEFMV.....NGVENPNRIDSPKKKKEKAP 363

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1892	100.0	363	21	AA198146
2	1892	100.0	363	22	AA198146
3	1463	77.3	369	21	AA198147
4	1463	77.3	369	22	AA198147
5	1463	77.3	369	22	AA198147
6	1463	77.3	369	22	AA198147
7	1134.5	60.0	416	21	AA198147
8	522	48.7	304	20	AA198147
9	739	38.1	384	22	AA198147
10	566	29.9	368	22	AA198147
11	566	29.9	368	22	AA198147

12	566	29.9	1575	22	AB167362	Drosophila melanog
13	414.5	21.9	125	21	AA198189	Human secreted pro
14	148.5	7.8	411	17	AA1986810	Saccharomyces cere
15	138.5	7.3	394	22	AA1980876	Human LAPH-1 prote
16	138.5	7.3	394	22	AA1980876	Human protein sequ
17	138.5	7.3	394	22	AA1980876	Human protein sequ
18	133.5	7.1	191	21	AA1984489	Human secreted pro
19	117	6.2	488	22	AA1982028	Human polypeptide
20	117	6.2	536	22	AA1980876	Human protein sequ
21	115.5	6.1	286	21	AA19844196	Arabidopsis thalia
22	114.5	6.1	280	21	AA19844196	Arabidopsis thalia
23	111.5	5.9	158	22	AA1981576	Novel signal trans
24	111.5	5.9	181	22	AA1981701	Novel signal trans
25	111	5.9	312	22	AA1982224	Human olfactory re
26	111	5.9	312	22	AA1982224	Human OR-like poly
27	110	5.8	380	20	AA1980877	Human LAPH-2 prote
28	110	5.8	380	22	AA1980877	Human hydrophobic
29	108	5.7	311	22	AA1980782	Human apoptosis pr
30	106.5	5.6	266	21	AA1983806	Human cancer assoc
31	105.5	5.6	310	21	AA1981779	Arabidopsis thalia
32	101.5	5.4	150	21	AA1984938	Human secreted pro
33	98	5.2	232	22	AB167362	Drosophila melanog
34	97.5	5.2	313	21	AA1983404	zebra mays protein f
35	97.5	5.2	397	20	AA1983316	Chlamydia trachoma
36	97.5	5.2	533	22	AA19840417	Human polypeptide
37	96.5	5.1	431	22	AA19830056	Novel human diagn
38	96	5.1	254	21	AA1981780	Arabidopsis thalia
39	93	4.9	313	21	AA1980877	Human G protein-co
40	93	4.9	313	22	AA1980877	Human olfactory re
41	92	4.9	313	22	AA1980877	Human olfactory re
42	91	4.8	307	22	AA1982203	Human olfactory re
43	91	4.8	315	22	AA1982870	Human olfactory re
44	91	4.8	324	22	AA1981539	Human olfactory re
45	90	4.8	309	22	AA19810312	G-protein coupled

ALIGNMENTS

RESULT 1	
AA198146	
ID	AA198146 standard; Protein; 363 AA.
AC	AA198146;
DT	22-AUG-2000 (first entry)
XX	Rat WAR-1 amino acid sequence.
DE	Endoplasmic reticulum; WAR-1; cancer cell proliferation inhibitor;
KW	diagnosis; cancer; sarcoma; rat.
XX	Rattus norvegicus.
XX	WO200022123-A1.
PN	20-APR-2000.
XX	13-OCT-1999; 99WO-JP05631.
PF	13-OCT-1998; 98JP-0290711.
XX	(SUMO) SUMITOMO PHARM CO LTD.
PA	Tohdoh N, Yoshima T, Komiya K, Tojo S, Nemoto K, Ishikawa H;
XX	Okuyama H;
PI	WPI: 2000-317980/27.
XX	N-PSDB; AAA38012.
DR	Endoplasmic reticulum protein WAR-1 which inhibits cancer cell
XX	proliferation for use in treatment and diagnosis of cancer including
PT	sarcomas of high malignancy -

XX Claim 1; Fig 2; 89pp; Japanese.
 XX This sequence represents an endoplasmic reticulum protein (WAR-1) amino
 CC acid sequence. The invention includes rat and human WAR-1 sequences,
 CC expression vectors containing the DNA, cells transformed with the
 CC expression vector, antibodies against WAR-1, and probes and primers which
 CC hybridise to the DNA encoding WAR-1. The WAR-1 protein inhibits the
 CC proliferation of cancer cells, and is used in the treatment and diagnosis
 CC of cancers including highly malignant sarcomas.
 XX Sequence 363 AA;

Query Match 100.0%; Score 1892; DB 21; Length 363;
 Best Local Similarity 100.0%; Pred. No. 3.8e-213;
 Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLRKKARNPPVLSHEPMQNHADMVSCVGMFFVLGLMFEFTAEMSVFTLTQHGYYVP 60
 Db 1 mglrkkarnppvlshemqnhadmvcvgmffvlgmfeftaemsvftltqhgvyvp 60
 QY 61 AEGLPSSGRTLYHYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRLQTKGQNKLINE 120
 Db 61 aeglpssgtrtlyhygvkdlatvffymlyvaaiihatiqeyvldklsrrlqtkgqnkline 120
 QY 121 AGQLSVFYIVSGIWMILASENCLSDPTLLWKSQPHNMFTFQMKFFYISQLAYWFHSFP 180
 Db 121 agqlsvfyivsgiwgmilasencldptllwksqphnmmtfkmkffysqlaywfhsfp 180
 QY 181 ELYFQKVRKQDIPGQLIYIGLHFIHGAYLLYLNHLGLLLMLHYAVELLSSVCSLLYF 240
 Db 181 elyfqvrkqdipgqliyiglhfihgayllylnhglglllmlhyavellssvcsillyf 240
 QY 241 GDERYQKGLSIPVIFISGRVLTIVSVTVGLHLAGTNRNGNALSGNVNLAATAVLVS 300
 Db 241 gderykqglslvpvifisgrvltivsvtvglhlagtnrngnalsgnvnvlaatavlvs 300
 QY 301 SSCSIQYIYITWTITVWLQWLEADANLHVCGRKRRSRKGTENGVENPNRIDSPPKKKE 360
 Db 301 sscsiqyitwtittvwlqrlwledanlhvcgrkrrsrkgtengvenpnridspppkke 360
 QY 361 KAP 363
 Db 361 kap 363

RESULT 2
 AAB70696
 ID AAB70696 standard; Protein; 363 AA.
 AC AAB70696;
 XX
 DT 17-MAY-2001 (first entry)
 DE
 XX
 XX Rat WAR-1 protein sequence SEQ ID NO:2.
 KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
 KW endoplasmic reticulum membrane transportation; secretory protein;
 KW cell membrane protein; cytosolic; CNS active; antiallergic; cancer;
 KW antirheumatic; nervous system disorder; immune disorder; allergy;
 KW rheumatism; skeletal disorder.
 XX
 OS Rattus sp.
 XX
 XX WO200114582-A1.
 PN
 XX
 PD 01-MAR-2001.
 XX
 XX 17-AUG-2000; 2000WO-JP05488.
 PF
 XX
 XX 20-AUG-1999; 99JP-0234764.
 PR
 XX

PA (SUMU) SUMITOMO PHARM CO LTD.
 XX Tohdoh N, Okuyama H, Inamura M, Ishikawa H, Nemoto K;
 PI WPI: 2001-202940/20.
 XX N-PSDB; AAF74781.
 DR Transformation of a cell with separate vectors expressing the sense and
 XX antisense strands of WAR-1 DNA for screening secretory and membrane
 PT proteins expressed by the cell
 PT
 PS Claim 3; Page 60-62; 79pp; Japanese.
 XX The present invention describes a screening method for secretory and
 CC membrane proteins consisting of transformation of a cell with separate
 CC expression vectors for the sense and antisense RNA of DNA encoding an
 CC endoplasmic reticulum membrane protein participating in endoplasmic
 CC reticulum transport of proteins. Also described are: (1) secretory and
 CC cell membrane proteins identified by the screening method; (2) drug
 CC compositions containing these proteins; (3) host cells transformed by
 CC the separate expression vectors of the method; and (4) the preparation
 CC of secretory and cell membrane proteins by culture of the transformants.
 CC The method can be used for the identification and preparation of
 CC proteins for use in the treatment and prevention of diseases such as
 CC cancer, disorders of the nervous system, immune disorders (including
 CC allergies and rheumatism) and skeletal disorders. The present sequence
 CC represents a specifically claimed rat WAR-1 protein from the present
 XX invention.
 SQ Sequence 363 AA;

Query Match 100.0%; Score 1892; DB 22; Length 363;
 Best Local Similarity 100.0%; Pred. No. 3.8e-213;
 Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLRKKARNPPVLSHEPMQNHADMVSCVGMFFVLGLMFEFTAEMSVFTLTQHGYYVP 60
 Db 1 mglrkkarnppvlshemqnhadmvcvgmffvlgmfeftaemsvftltqhgvyvp 60
 QY 61 AEGLPSSGRTLYHYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRLQTKGQNKLINE 120
 Db 61 aeglpssgtrtlyhygvkdlatvffymlyvaaiihatiqeyvldklsrrlqtkgqnkline 120
 QY 121 AGQLSVFYIVSGIWMILASENCLSDPTLLWKSQPHNMFTFQMKFFYISQLAYWFHSFP 180
 Db 121 agqlsvfyivsgiwgmilasencldptllwksqphnmmtfkmkffysqlaywfhsfp 180
 QY 181 ELYFQKVRKQDIPGQLIYIGLHFIHGAYLLYLNHLGLLLMLHYAVELLSSVCSLLYF 240
 Db 181 elyfqvrkqdipgqliyiglhfihgayllylnhglglllmlhyavellssvcsillyf 240
 QY 241 GDERYQKGLSIPVIFISGRVLTIVSVTVGLHLAGTNRNGNALSGNVNLAATAVLVS 300
 Db 241 gderykqglslvpvifisgrvltivsvtvglhlagtnrngnalsgnvnvlaatavlvs 300
 QY 301 SSCSIQYIYITWTITVWLQWLEADANLHVCGRKRRSRKGTENGVENPNRIDSPPKKKE 360
 Db 301 sscsiqyitwtittvwlqrlwledanlhvcgrkrrsrkgtengvenpnridspppkke 360
 QY 361 KAP 363
 Db 361 kap 363

RESULT 3
 AAY98147
 ID AAY98147 standard; Protein; 369 AA.
 XX
 AC AAY98147;
 XX
 DT 22-AUG-2000 (first entry)
 XX

DE Human WAR-1 amino acid sequence.
XX
KW Endoplasmic reticulum; WAR-1; cancer cell proliferation inhibitor;
KW diagnosis; cancer; sarcoma; human.
OS Homo sapiens.
XX
PN WO200022123-A1.
XX
PD 20-APR-2000.
XX
PE 13-OCT-1999; 99WO-JP05631.
XX
PR 13-OCT-1998; 98JP-0290711.
XX
PA (SUMI) SUMITOMO PHARM CO LTD.
XX
PI Tachihara N, Yoshida T, Komiya K, Tojo S, Nemoto K, Ishikawa H;
PI Okuyama H;
DR WPI; 2000-317980/27.
DR N-PSDB; AAA38013.
XX
XX Endoplasmic reticulum protein WAR-1 which inhibits cancer cell
PT proliferation for use in treatment and diagnosis of cancer including
PT sarcomas of high malignancy
PT
XX
PS Claim 1; Fig 2; 89pp; Japanese.
XX
CC This sequence represents an endoplasmic reticulum protein (WAR-1) amino
CC acid sequence. The invention includes rat and human WAR-1 sequences,
CC expression vectors containing the DNA, cells transformed with the
CC expression vector, antibodies against WAR-1, and probes and primers which
CC hybridize to the DNA encoding WAR-1. The WAR-1 protein inhibits the
CC proliferation of cancer cells, and is used in the treatment and diagnosis
CC of cancers including highly malignant sarcomas.
XX
SQ Sequence 369 AA;

Query Match 77.3%; Score 1463; DB 21; Length 369;
Best Local Similarity 75.8%; Pred. No. 9.6e-163;
Matches 279; Conservative 43; Mismatches 40; Indels 6; Gaps 5;
QY 1 MGLRKKRNARPPVLSHEFMVONHADMSCVGMFFVLGMFEFTAEMSIIVFLTLQHGVPVP 60
DB 1 mglrkkstknppvlsqefilqnhadivscvgmffilgvlfgtaeasivfltlqnsavp 60
QY 61 -AEGLPSSGRTLYHYGVKDLATVFFYMLVAIIIIHATIQEYVLDKLSRRLOLTKGQKNLN 119
DB 61 aaeqatgsksllyygvkdlatvffymvliiiahatiqeyvldklnrmtftkakqkn 120
QY 120 EAQQLSVFYIVSGIWMIIASNCISDPLLLWKSQPHNMFTQMKFFYISQLAYWPHSF 179
DB 121 esqgfsyvfiffciglwtfflilsencisdpdlwkarphsmntfmgkffysqlaywfhaf 180
QY 180 PELYFQVRKQDIPGQIYVIGLHFGGAYLLYLNHLGLLMLHVAVELLSVCSLLY 239
DB 181 pelyfqtakqdlprqlvylglnhflhtgagyllylnhlgilvlyhvfellsmcgilfy 240
QY 240 FGDERYQKGLSLWPIVIFISGRVLTIVSVVTVGLHLAGT--NRNGNLSGNVNVLAAKIIV 298
DB 241 fsdekyqkglslwaivfilgrlvtllsvltvghlagsqnrpdaltgnvnlakiaiv 300
QY 299 LSSCSIQYITWTLTVWQLWLEDANLHV-CGRKRSR-SRKGTEN--GVENPNRIDS 354
DB 301 lsscsitqayvtwnltlwlqrvvedsnlqascmkkrsskrrtengvgvetsnrvc 360
QY 355 PPKKKEKA 362
DB 361 ppkkrkes 368

RESULT 4
ABG12234
ID ABG12234 standard; Protein; 369 AA.
XX
AC ABG12234;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #12225.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS76421.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 42593; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 369 AA;

Query Match 77.3%; Score 1463; DB 22; Length 369;
Best Local Similarity 75.8%; Pred. No. 9.6e-163;
Matches 279; Conservative 43; Mismatches 40; Indels 6; Gaps 5;
QY 1 MGLRKKRNARPPVLSHEFMVONHADMSCVGMFFVLGMFEFTAEMSIIVFLTLQHGVPVP 60
DB 1 mglrkkstknppvlsqefilqnhadivscvgmffilgvlfgtaeasivfltlqnsavp 60
QY 61 -AEGLPSSGRTLYHYGVKDLATVFFYMLVAIIIIHATIQEYVLDKLSRRLOLTKGQKNLN 119
DB 119

Db 61 aaeqatgsksllyygvkdlatvffymlvaiiiahatiqeyvldkinkrmqftkakqkfn 120
QY 120 EAGQLSVFYIVSGIWMGMIILASENCLSDPTLLWKSPHNMFTQMKFFYISOLAYWFHSF 179
Db 121 esgqfsvfyfscilwgtfllisencldptllwkarphsmmtfcmkffysqlaywfhaf 180
QY 180 PELYFQVRKQDIPGQLIYIGLHFLHIGGAYLLYLNHLGLLMLHYAVELLSSVCSILLY 239
Db 181 pelyfqtktkqdiqrlyvlgihlhitgalyllnhlgllllylvhyfvelshmcglfy 240
QY 240 FGDERYQKGLSLMPVIFISGRVLTLIVSVTVGLHLAGT-NRNGNALSGNVNVLAAKIAV 298
Db 241 fsdekyqkgsilwailgrivtlivsvltvgfhlagsgnrdaltgngvnvlaakiav 300
QY 299 LSSSCSIQVYITWTLTWLQWLEADANLHV-CGRKRSR-SRKGTEN--GVENPNRIDS 354
Db 301 lssscsiqyvtwnltlwlqrwedsniqascmkkrsskrktengvgvetsnrvc 360
QY 355 PPKKKEKA 362
Db 361 ppkrkeks 368
RESULT 5
ID AAB70695 standard; Protein; 369 AA.
XX
AC AAB70695;
XX
DT 17-MAY-2001 (first entry)
XX
DE Human WAR-1 protein sequence SEQ ID NO:1.
KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
KW endoplasmic reticulum membrane transportation; secretory protein;
KW cell membrane protein; cytosolic; CNS active; antiallergic; cancer;
KW antirheumatic; nervous system disorder; immune disorder; allergy;
KW rheumatism; skeletal disorder.
OS Homo sapiens.
XX
PN WO200114582-A1.
XX
PD 01-MAR-2001.
XX
PF 17-AUG-2000; 2000WO-JP05488.
XX
PR 20-AUG-1999; 99JP-0234764.
XX
PA (SUMI) SUMITOMO PHARM CO LTD.
XX
PI Tohdoh N, Okuyama H, Imamura M, Ishikawa H, Nemoto K;
XX
DR WPI; 2001-202940/20.
DR N-PSDB; AAF74780, AAF74782.
XX
PT Transformation of a cell with separate vectors expressing the sense and
PT antisense strands of WAR-1 DNA for screening secretory and membrane
PT proteins expressed by the cell -
XX
PS Claim 2; Page 58-60; 79pp; Japanese.
XX
CC The present invention describes a screening method for secretory and
CC membrane proteins consisting of transformation of a cell with separate
CC expression vectors for the sense and antisense RNA of DNA encoding an
CC endoplasmic reticulum membrane protein participating in endoplasmic
CC reticulum transport of proteins. Also described are: (1) secretory and
CC cell membrane proteins identified by the screening method; (2) drug
CC compositions containing these proteins; (3) host cells transformed by
CC the separate expression vectors of the method; and (4) the preparation
CC of secretory and cell membrane proteins by culture of the transformants.
CC The method can be used for the identification and preparation of
CC proteins for use in the treatment and prevention of diseases such as

CC cancer, disorders of the nervous system, immune disorders (including
CC allergies and rheumatism) and skeletal disorders. The present sequence
CC represents a specifically claimed human WAR-1 protein from the present
CC invention.
XX
SQ Sequence 369 AA;
Query Match 77.3%; Score 1463; DB 22; Length 369;
Best Local Similarity 75.8%; Pred. No. 9,6e-163;
Matches 279; Conservative 43; Mismatches 103; Indels 6; Gaps 5;
QY 1 MGLRKKNARNPPVLSHEFMVONHADVSCVGMFVLCIMPEGTAEISMIVTLTQHGVVWP 60
Db 1 mglrkstknppvlsqefilqhnhadivscgmffllglviegtaeasivfltiqhsavp 60
QY 61 -ABGLPSGRTLYHYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRLQLTGKQKLN 119
Db 61 aaeqatgsksllyygvkdlatvffymlvaiiiahatiqeyvldkinkrmqftkakqkfn 120
QY 120 EAGQLSVFYIVSGIWMGMIILASENCLSDPTLLWKSPHNMFTQMKFFYISOLAYWFHSF 179
Db 121 esgqfsvfyfscilwgtfllisencldptllwkarphsmmtfcmkffysqlaywfhaf 180
QY 180 PELYFQVRKQDIPGQLIYIGLHFLHIGGAYLLYLNHLGLLMLHYAVELLSSVCSILLY 239
Db 181 pelyfqtktkqdiqrlyvlgihlhitgalyllnhlgllllylvhyfvelshmcglfy 240
QY 240 FGDERYQKGLSLMPVIFISGRVLTLIVSVTVGLHLAGT-NRNGNALSGNVNVLAAKIAV 298
Db 241 fsdekyqkgsilwailgrivtlivsvltvgfhlagsgnrdaltgngvnvlaakiav 300
QY 299 LSSSCSIQVYITWTLTWLQWLEADANLHV-CGRKRSR-SRKGTEN--GVENPNRIDS 354
Db 301 lssscsiqyvtwnltlwlqrwedsniqascmkkrsskrktengvgvetsnrvc 360
QY 355 PPKKKEKA 362
Db 361 ppkrkeks 368
RESULT 6
ID AAM93265 standard; Protein; 369 AA.
XX
AC AAM93265;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 2725.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
XX
PR 11-JAN-2000; 2000JP-0118774.
XX
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524355/58.
DR N-PSDB; AAK94181.
XX

PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
XX Claim 8; SEQ ID NO 2725; 1380pp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesizing the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is a polypeptide
XX encoded by a full length human cDNA of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.
SQ Sequence 369 AA;

Query Match 77.3%; Score 1462; DB 22; Length 369;
Best Local Similarity 75.3%; Pred. No. 1.3e-162;
Matches 278; Conservative 44; Mismatches 40; Indels 6; Gaps 5;
QY 1 MGLRKNARNPPVLSHEPVMONHADMVSCVGMFFVLGLMEFGTAEMSIIVTLQHGYYVVP 60
Db 1 mglrkstknpvlsqeflqhphadivscvgmffllglvfgtaeasivfltlqhsavp 60
QY 61 -AEGPSGRTLYHYGVKDLATVFFMLVAIIHATIQEYVLDKLSRLQLTKGKQNKLN 119
Db 61 aaeqatgskslvygvgkdlatvffmlvaliiahqeyvldklnrmqtkakqkfn 120
QY 120 EAGLSVFYIVSGTWGMIIILASENCLSDPTLLWKSQPHNMTFQMKFFYSOLAYWFHSF 179
Db 121 esgqfsyffscslwgffllsencsdptllwkarphsmntgmffkffysqlaywfhaf 180
QY 180 PELYFQVRKQDIPGQIYIYGLHFGGAYLLYLNHGLMLLMLHAYVELLSVCSLLY 239
Db 181 pelyfqtktkqdlprqlvyglhfhigtayllnhglvllvhyfvelshmcglfy 240
QY 240 FGDERYQKGLSWPFIPIISGRVTLIVSVTVGLHLAGT-NRNGNALSGNVNVAAKIAV 298
Db 241 fsdekygkglswaivfllgrvltvsvtgfhlagsgnmpdaltgnvnvaakiav 300
QY 299 LSSSCSTQVYITWTLLTVMLQRMLEDANLHV-CGRKRSR-SRKGTE--GVENPNRIDS 354
Db 301 lssscitqayvtwnltlwlgrwvdsnlqascmkkrsskkrtegnvgvetsnrvc 360
QY 355 PPRKKEKA 362
Db 361 pprkkes 368

RESULT 7
AAB43601
ID AAB43601 standard; Protein; 416 AA.
XX
AC AAB43601;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human cancer associated protein sequence SEQ ID NO:1046.
XX
XX Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
KW anti-diabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW anti-inflammatory; antithyroid; anti-allergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; neutropenic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;

KW neurological disease; drug screening.
XX Homo sapiens.
XX
XX W0200055350-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05882.
XX
XX 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX WPI; 2000-587533/55.
XX N-PSDB; AAC77810.
XX
XX Novel isolated nucleic acids comprising sequences encoding peptides
XX useful for treating or diagnosing e.g. cancer -
XX
XX Claim 11; Page 1634-1636; 2352pp; English.
XX
XX AAC77607 to AAC78448 encode the human cancer associated proteins given
XX in AAB43398 to AAB44239. The proteins can have activities based on the
XX tissues and cells the genes are expressed in. Example of activities
XX include: cytostatic; proliferative; vulnerable; immunomodulator;
XX anti-diabetic; antiasthmatic; antirheumatic; antiarthritic;
XX dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
XX neutropenic; vasotropic; antipsoriatic and antiangiogenic. The
XX polynucleotides and polypeptides can be used for preventing, treating or
XX ameliorating medical conditions and diagnosing pathological conditions.
XX polynucleotides, polypeptides, antibodies, agonists and antagonists from
XX the present invention may be used to treat immune disorders by activating
XX or inhibiting the proliferation, differentiation or mobilisation of
XX immune cells, to treat disorders of haematopoietic cells, autoimmune
XX disorders, allergic reactions, graft versus host disease and organ
XX rejection, modulate haemostatic or thrombolytic activity, modulate
XX inflammation, cancers, cardiovascular disorders, neurological disease and
XX bacterial or viral infections. The peptides, nucleotides, antibodies,
XX agonists and antagonists may be also be used in drug screens. AAC78449 to
XX AAC78457 and AAB44240 represent sequences used in the exemplification of
XX the present invention.
XX
XX Sequence 416 AA;

Query Match 60.0%; Score 1134.5; DB 21; Length 416;
Best Local Similarity 58.7%; Pred. No. 4.5e-124;
Matches 219; Conservative 61; Mismatches 82; Indels 11; Gaps 4;
QY 1 MGLRKNARNPPVLSHEPVMONHADMVSCVGMFFVLGLMEFGTAEMSIIVTLQHGYYVVP 60
Db 43 mairkstkspvlshefvlqnhadivscvgmffllglvfgtaeasivfltlqynvtlp 102
QY 61 A-EGLPSGRTLYHYGVKDLATVFFMLVAIIHATIQEYVLDKLSRLQLTKGKQNKLN 119
Db 103 ateeqatesvslvygvgkdlatvffmlvaliiahqeyvldklnrmhsktkhskfn 162
QY 120 EAGLSVFYIVSGTWGMIIILASENCLSDPTLLWKSQPHNMTFQMKFFYSOLAYWFHSF 179
Db 163 esgqlsafylfacwgtfllisenysdptllwrayphnmtfkmffysqlaywfhaf 222
QY 180 PELYFQVRKQDIPGQIYIYGLHFGGAYLLYLNHGLMLLMLHAYVELLSVCSLLY 239
Db 223 pelyfqtktkqdlprqlvyglhfhigtayllnhglvllvhyfvefifhsrlfy 282
QY 240 FGDERYQKGLSWPFIPIISGRVTLIVSVTVGLHLAGT-NRNGNALSGNVNVAAKIAV 298
Db 283 fsnekygkglswaivfllgrvltvsvtgfhlagsgnmpdaltgnvnvaakiav 342

```
QY 299 LSSSCSIQYITWTLTWTWQLRWLELDANLHVCGRRK-----SRSRKGTENGVE---NP 349
Db 343 lasicvtqatmmwkwfinfoqrrrehsafapavkkpvtvktgrsskktgengvngltls 402
QY 350 NRIDSPPKKEKA 362
Db 403 nvadspnkkks 415

RESULT 8
AAY48434
ID AAY48434 standard; Protein; 304 AA.
XX
AC AAY48434;
XX
DT 08-DEC-1999 (first entry)
XX
DE Human prostate cancer-associated protein 131.
KW Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;
KW cancer; tissue specificity; human.
XX
OS Homo sapiens.
XX
PN DE19811194-A1.
XX
PD 16-SEP-1999.
XX
PF 10-MAR-1998; 98DE-1011194.
XX
PR 10-MAR-1998; 98DE-1011194.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosenthal A;
DR WPI; 1999-519629/44.
DR N-PSDB; AA233531.
XX
XX New nucleic acid expressed at high level in normal prostatic tissue and
PT encoded polypeptides, used to treat cancer and screen for therapeutic
PT agents
XX
PS Claim 25; 174; 194pp; German.
XX
CC This invention describes novel nucleic acid sequences (A) that are
CC expressed at high level in normal prostatic tissue. Polypeptides (I)
CC encoded by (A) are used: (a) for identifying agents for treatment of
CC prostatic cancer and (b) for therapy of prostate cancer, optionally
CC where expressed by gene therapy methods. (A) is also used to isolate
CC full-length genes (for gene therapy) and for recombinant production of
CC (I), which can be used to raise specific antibodies. (A) are identified
CC by assembly of ESTs (expressed sequence tags) before these are analyzed
CC for expression pattern (tissue specificity). This approach eliminates
CC many of the false results, as regards tissue specificity, associated
CC with known methods that use single (usually short) ESTs. AAY48304-Y48456
CC represent peptides encoded by the expressed sequence tags described in
CC the method of the invention.
XX
SQ Sequence 304 AA;

Query Match 48.7%; Score 922; DB 20; Length 304;
Best Local Similarity 58.1%; Pred. No. 2.5e-99;
Matches 176; Conservative 48; Mismatches 69; Indels 10; Gaps 3;

QY 70 TLYHYGVKDLATVFFVFWLVAIIHATIQEYVDLKLRRRLQTLTKGKONKLNKAGQLSVFYI 129
Db 1 slyyygirdlatvffymlyvaihaviqeymldklnrmhfskthkskfnesgqlsafyl 60
QY 130 VSGIWMILLAGENCLSDPTLWKSOPHNMTFQMKFFYISOLAYWFHSPFELYFQKVRK 189
Db 61 facvwtflisenyisdptllwrayphnmfymkffiyisqlaywihafpelyfdtkk 120
```

```
QY 190 QDIPGQLIYIGLHDFHIGGAYLLYLNHLGLLLMLHYAVELLSSVCSLLYFGDERYQKGL 249
Db 121 ediprqlviygllyfhiagayllnlnhglvllvlhyfveflhlsrlfyfsnekyqkgf 180
QY 250 SLWPIFVIFISGRLVTLIVSVTVGLHLA-GTNRNGNALSGNVNLAAKTAVLSSSCSIQYV 308
Db 181 slwavlflvgrltilsvitvgfglaraenqkldfstgnfnvlavriaviasicvtqaf 240
QY 309 ITWTLTWTWQLRWLELDANLHVCGRRK-----SRSRKGTENGVE---NPNRIDSPPKKK 359
Db 241 mmwkwfinfoqrrrehsafapavkkpvtvktgrsskktgengvngltlsnvadspnkk 300
QY 360 EKA 362
Db 301 eks 303

RESULT 9
AAM95708
ID AAM95708 standard; Protein; 384 AA.
XX
AC AAM95708;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen SEQ ID NO: 4366.
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 28-JUN-2000; 2000US-0209467.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
```


PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-0259678.
XX
XX

(HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-465570/50.

DR N-PSDB; AAL01678.

XX Isolated nucleic acid molecule encoding a reproductive system antigen -
PT is used in preventing, treating or ameliorating a medical condition -

XX Claim 11; SEQ ID NO 4366; 1297pp + Sequence Listing; English.

XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a protein of the invention.

XX Sequence 384 AA;

Query Match 39.1%; Score 739; DB 22; Length 384;

Best Local Similarity 41.8%; Pred. No. 1.le-77;

Matches 156; Conservative 62; Mismatches 135; Indels 20; Gaps 5;

QY 5 KKNARNPPVLSEHFVNQADWVSCVGMFFVLGLMFEQTAEMSVLFTLQHGCVVVPBGL 64

Db 18 rrrtkysplfsqefvihnhdigfcvlvligImfxvtaktafllpqynisvpt--- 74

QY 65 PSGSRTL-YHGVKDLATVFFYMLVAIIHATIQEIVLDKLSRLQLTKGKONKLINEAG 123

Db 75 -adsetvhygpkdlvtilfifitilhavvqeyildkiskrlhskvkhskfnesgq 133

QY 124 LSVFYIVSGIWMITILASENCLSDPTLLWKSOPHNMTFQMKFFYISQLAYWFHSFPPELY 183

Db 134 lvvfhtsvlwcfcyvvvtegylnprslwedvphvhpqvkfycqlaylhalpely 193

QY 184 FQVVRKQDIPQQLIYIGLHLFHHGAYLLYLNHLGLLMLHLHYAVEVLSVSSVCLLYFQDE 243

Db 194 fqvrkeeiprqlqyiclyvhiagaylnlsrlgllllllyqstefifhcxrifyfade 253

QY 244 RYQKGLSLWPIVYFSGRLVTLIVSVTVGLHLAG-TNRNGNALSGNVNLAAKTAVLSSS 302

Db 254 nneklfsawaaxfgvtrflfllavlaigfglarmenqafdpkgnfnllcrilcvllv 313


```
Query Match          29.9%; Score 566; DB 22; Length 368;
Best Local Similarity 37.0%; Pred. No. 2.2e-57;
Matches 136; Conservative 64; Mismatches 130; Indels 38; Gaps 9;

QY 4 RKNARNPPVLSHEFMVQNHADVMSCVGMFFVLGLMFEGETAEMSIVFLTLQHG--VWPA 61
DB 9 rKtsnknppilshfviqnhadiiscvamfvvglmnestaafasafisihhnvsgebps 68
QY 62 EGLPSGSRTLYHYGVKDLATVFFVYMLVAIIHATIOEVVLDKLSRRLQLTGKQNKLEA 121
DB 69 reqpygkptyiagikdycaiffytclimhaiqefvdkiskkhlskfklafrnes 128
QY 122 GQLSVFYIVSGIWMGIILASENCLSDPTLLWKSQPHNMFTQMKFFYISQLAYWFHSPFE 181
DB 129 gqlvafyllsfvqahvllkegylgvqqlwegfdpmsflhkfyfvvqlayylhmlpe 188
QY 182 LYQKVR-KQDIPGQLIYIGLHFIHGG-----AYLLYNHLGLLLMLHYAVELLSSV 234
DB 189 lyfqiktkkeeqqpkivh-----sigsftliivlaytisfqlalvltlthfseillshv 242
QY 235 CSLL--YFGDERYOKGLSLWPIVIFISGRVLTIVSVTVGLHAGTNRNGNALSQNVNL 292
DB 243 fqligvfdreerlaklrsvnnnavflirfatsvigtlyyigv-rsllaiggli--- 298
QY 293 AAKIAYLSSSCSIQVYITWTTLTVWLQRLWLEDANLHVCGRRKRSRSGKTENGVENPNRI 352
DB 299 -----alqgylvfsfiteqlrakreakk----eakreaklalktkkpktpk-- 341
QY 353 DSPPKKKE 360
DB 342 dkvkrkke 349

RESULT 12
ABB67362 ID ABB67362 standard; Protein; 1575 AA.
XX AC ABB67362;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 28878.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL11465.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 28878; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
```

```
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB1616176-ABL30511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1575 AA;
SQ
```

```
Query Match          29.9%; Score 566; DB 22; Length 1575;
Best Local Similarity 37.0%; Pred. No. 2e-56;
Matches 136; Conservative 64; Mismatches 130; Indels 38; Gaps 9;

QY 4 RKNARNPPVLSHEFMVQNHADVMSCVGMFFVLGLMFEGETAEMSIVFLTLQHG--VWPA 61
DB 1216 rKtsnknppilshfviqnhadiiscvamfvvglmnestaafasafisihhnvsgebps 1275
QY 62 EGLPSGSRTLYHYGVKDLATVFFVYMLVAIIHATIOEVVLDKLSRRLQLTGKQNKLEA 121
DB 1276 reqpygkptyiagikdycaiffytclimhaiqefvdkiskkhlskfklafrnes 1335
QY 122 GQLSVFYIVSGIWMGIILASENCLSDPTLLWKSQPHNMFTQMKFFYISQLAYWFHSPFE 181
DB 1336 gqlvafyllsfvqahvllkegylgvqqlwegfdpmsflhkfyfvvqlayylhmlpe 1395
QY 182 LYQKVR-KQDIPGQLIYIGLHFIHGG-----AYLLYNHLGLLLMLHYAVELLSSV 234
DB 1396 lyfqiktkkeeqqpkivh-----sigsftliivlaytisfqlalvltlthfseillshv 1449
QY 235 CSLL--YFGDERYOKGLSLWPIVIFISGRVLTIVSVTVGLHAGTNRNGNALSQNVNL 292
DB 1450 fqligvfdreerlaklrsvnnnavflirfatsvigtlyyigv-rsllaiggli--- 1505
QY 293 AAKIAYLSSSCSIQVYITWTTLTVWLQRLWLEDANLHVCGRRKRSRSGKTENGVENPNRI 352
DB 1506 -----alqgylvfsfiteqlrakreakk----eakreaklalktkkpktpk-- 1548
QY 353 DSPPKKKE 360
DB 1549 dkvkrkke 1556
```

```
RESULT 13
AAG00189 ID AAG00189 standard; Protein; 125 AA.
XX AC AAG00189;
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein, SEQ ID NO: 4270.
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping.
XX OS Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX PR 21-FEB-2000; 2000EP-0200610.
XX PR 26-FEB-1999; 99US-0122487.
XX PA (GEST ) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
```

DR WPI; 2000-500381/45.
 XX N-PSDB; AAC00195.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 XX
 PS Claim 13; SEQ ID 4270; 71pp + CD-ROM; English.
 XX
 XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 XX
 SQ Sequence 125 AA;

Query Match 21.9%; Score 414.5; DB 21; Length 125;
 Best Local Similarity 64.0%; Pred. No. 2.8e-40;
 Matches 80; Conservative 24; Mismatches 20; Indels 1; Gaps 1;
 Qy 1 MGLRKKARNPPVLSHEFMVQNHADMSCVGMFFVLGLMFEGETAEMSIVFLTLQHGYYVVP 60
 Db 1 mairkstkspvlshefvqlphadivscvamvflglmlfteitakasiifvtlqynvltlp 60
 Qy 61 A-EGLPSCSRTRYHYGVKDLATVFYMLVAIIHATTOEIVLDKLSRRLQLTGKQKNLN 119
 Db 61 ateeqatesvlyyygikdiatvffymvlvaifhavigeymldkinrmxfsktxhskfn 120
 Qy 120 EAGQL 124
 Db 121 esgql 125

RESULT 14
 AAR86810
 ID AAR86810 standard; Protein; 411 AA.
 XX
 AC AAR86810;
 XX
 DT 28-MAY-1997 (first entry)
 XX
 DE Saccharomyces cerevisiae LAG1 protein.
 XX
 KW Saccharomyces cerevisiae; LAG1; life-span limiting domain;
 KW life-span extending domain; stress tolerance; longevity;
 KW recombinant protein production.
 XX

Saccharomyces cerevisiae.
 XX
 XX Key Location/Qualifiers
 FT Domain 1..190
 FT /note= "life-span limiting domain"
 FT 194..411
 FT /note= "life-span extending domain"
 XX

PN WO9533834-A1.
 XX
 XX 14-DEC-1995.
 PD
 XX 02-JUN-1995; 95WO-US06725.
 PF
 XX 08-NOV-1994; 94US-0336031.
 PR
 XX 03-JUN-1994; 94US-0253875.
 PR

XX
 PA (RESE) RESEARCH CORP TECHNOLOGIES INC.
 XX Jazwinski SM;
 PI
 XX WPI; 1996-040238/04.
 DR N-PSDB; AAT07263.
 XX
 XX Eukaryotic LAG1 gene and protein - controls longevity, stress
 PT tolerance and reproductive capacity of eukaryotic cells, for
 PT improved prodn. of recombinant proteins.
 XX
 XX Claim 23; Page 100-102; 154pp; English.
 XX
 XX The Saccharomyces cerevisiae LAG1 gene product is composed of a
 CC life-span limiting domain (see AAR86812) and a life-span extending
 CC domain (see AAR86811). Overexpression of the LAG1 gene in older
 CC cells has a rejuvenating effect, which not only increases
 CC cellular life span, but also reproductive capacity and cellular
 CC tolerance to stress factors such as starvation and low pH. These
 CC cells may be used for the production of recombinant proteins. By
 CC increasing the life span of recombinant cells, the need for
 CC overexpression of recombinant gene products is avoided, and
 CC therefore any subsequent adverse effects on the host cell.
 XX
 SQ Sequence 411 AA;

Query Match 7.8%; Score 148.5; DB 17; Length 411;
 Best Local Similarity 22.0%; Pred. No. 2.9e-08;
 Matches 68; Conservative 46; Mismatches 106; Indels 89; Gaps 11;
 Qy 26 MVSVCVGMFFVLGLMFEGETAEMSIVFLTLQHGYYVVPABGLPSSGSRTRYHYGVKDLATVFFY 85
 Db 91 lvcvysayflisgrntesnplhmfvaisyq-----vdgtds-----yakgikdlsfvyf 139
 Qy 86 MLVAIIHATTOEIVLDKLSRRLQLTGKQKNLNEAGQLSVFYI-VSGIWGMIL-ASEN 143
 Db 140 mifftflreflmdvvrpftvynvtsehrqkrmleqmaifcvgvsgpgfglymyhsdl 199
 Qy 144 CLSDPTLLWKSQPHNMFTFQMKFFYISQLAYW-----FHSFPE 181
 Db 200 wlfktkpmryrtypitnplfkifiygqaafwaqacvvlqllekprkdykeivfhivt 259
 Qy 182 L-----YFQVRKQDIPGQLIYI-----GLHLFH 205
 Db 260 llllwsyvfhtkm-----glaiyitmdvsdfllsktlaylnsvftpfvgflvf- 312
 Qy 206 IGGAYLLYLNHLGLLLMLHYAVELLSSVCSLLYFGDERYQKGLSLMPYIFISGRLVTLI 265
 Db 313 ----fwylrhnvnrilwslvtefrhgyvlnfatqgkwcwisl-plivfv-----ll 361
 Qy 266 VSVVTVCLH 274
 Db 362 aaiqlvnly 370

RESULT 15
 AAY00876
 ID AAY00876 standard; Protein; 394 AA.
 XX
 AC AAY00876;
 XX
 DT 21-MAY-1999 (first entry)
 XX
 DE Human LAPH-1 protein sequence.
 XX
 KW Longevity-assurance protein homologue; LAPH-1; human; cell proliferation;
 KW longevity-assurance protein homologue; LAPH-2; signal transduction;
 KW cell cycle regulation; apoptosis; cellular homeostatic pathway; aging;
 KW cancer; inflammation; autoimmune disease; infection;
 KW neurodegenerative disorder.
 XX

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 14:57:46 ; Search time 28.82 Seconds
(without alignments)
307.651 Million cell updates/sec

Title: US-09-807-470-2
Perfect score: 1892
Sequence: 1 MGLRKNARNPPVLSHEFMV.....NGVENPNRIDPPKKKEKAP 363

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/PT05_COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	148.5	7.8	411	2	US-08-336-031-2
2	148.5	7.8	411	2	US-08-902-853-7
3	148.5	7.8	411	5	PCT-US95-06725-2
4	138.5	7.3	394	2	US-08-902-853-1
5	125	6.6	387	2	US-08-902-853-6
6	110	5.8	380	2	US-08-902-853-3
7	89	4.7	649	4	US-08-800-291B-5
8	89	4.7	649	4	US-08-800-291B-6
9	89	4.7	650	4	US-08-800-291B-4
10	87	4.6	1476	4	US-09-256-703-2
11	87	4.6	1479	2	US-08-951-912-4
12	87	4.6	1479	4	US-09-174-077-4
13	87	4.6	1480	1	US-07-637-621-2
14	87	4.6	1480	1	US-08-136-742A-2
15	87	4.6	1480	1	US-08-135-809A-2
16	87	4.6	1480	1	US-08-466-886-17
17	87	4.6	1480	1	US-08-951-912-2
18	87	4.6	1480	2	US-08-951-912-6
19	87	4.6	1480	2	US-08-469-461-2
20	87	4.6	1480	2	US-08-469-461-4
21	87	4.6	1480	2	US-08-691-605-2
22	87	4.6	1480	2	US-08-455-552A-14
23	87	4.6	1480	3	US-07-890-609-2
24	87	4.6	1480	3	US-07-890-609-4
25	87	4.6	1480	3	US-09-248-026-2
26	87	4.6	1480	4	US-08-469-617-17
27	87	4.6	1480	4	US-08-681-838A-2

28	87	4.6	1480	4	US-08-681-838A-3	Sequence 3, Appli
29	87	4.6	1480	4	US-09-174-077-2	Sequence 2, Appli
30	87	4.6	1480	4	US-09-174-077-6	Sequence 6, Appli
31	87	4.6	1480	5	PCT-US93-11667-2	Sequence 2, Appli
32	87	4.6	1480	6	5240846-5	Patent No. 5240846
33	85	4.5	193	2	US-08-336-031-6	Sequence 6, Appli
34	85	4.5	193	5	PCT-US95-06725-6	Sequence 6, Appli
35	82	4.3	2183	3	US-08-746-111-5	Sequence 5, Appli
36	80	4.2	617	1	US-07-879-617A-11	Sequence 11, Appli
37	80	4.2	617	1	US-08-301-722A-3	Sequence 3, Appli
38	80	4.2	617	1	US-08-240-783B-3	Sequence 3, Appli
39	80	4.2	617	1	US-08-753-985-11	Sequence 11, Appli
40	80	4.2	617	3	US-09-084-813-3	Sequence 3, Appli
41	80	4.2	617	5	PCT-US92-09662-3	Sequence 3, Appli
42	79	4.2	326	4	US-08-986-768-2	Sequence 2, Appli
43	79	4.2	602	1	US-08-295-814E-2	Sequence 2, Appli
44	79	4.2	602	4	US-09-343-361-2	Sequence 2, Appli
45	79	4.2	602	5	PCT-US93-01959-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-336-031-2
; Sequence 2, Application US/08336031
; Patent No. 5817782
; GENERAL INFORMATION:
; APPLICANT: Jazwinski, S. M.
; TITLE OF INVENTION: LAG1: A GENE FOR INCREASING THE
; TITLE OF INVENTION: LONGEVITY OF EURARYOTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,031
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/253,875
; FILING DATE: 03-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 9303Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-336-031-2

Query Match 7.8%; Score 148.5; DB 2; Length 411;
Best Local Similarity 22.0%; Pred. No. 1.1e-08;
Matches 68; Conservative 46; Mismatches 106; Indels 89; Gaps 11;
QY 26 MVSCVGMFFVLGLMFEGETAEMSIIVFLTLQGVVVVPAEGLPSGSRITLYHYGVKDLATVFFV 85

Db 91 LVCVTSAYFLSGNRTESNPLHMFVAISYQ-----VDGTDS-----YAKGKDLVSFVFY 139
QY 86 MLVAIIIIHATIQEYVLDKLSRRLOITKQKNEAGOLSVFYI-VSGIWMIL-ASEN 143
Db 140 MIFTFLEFLMDVVIRPFTYVLTNSHQRKMLEQMAIYFCVSPFGLYIMYHSDL 199
QY 144 CLSDPTLLWKSOPHNMFTQMKFFYISQAYW-----FHSPE 181
Db 200 WLFKTKPMYRTPVTNPFLEKIFVLGQAAEWAQACVVLQLEKPRKDYKELVFHHVT 259
QY 182 L-----YFQVKRQDIPGQIYI-----GLHLFH 205
Db 260 LLLIWSYVVFHFTKM-----GLAIYITMDVSDFFLSKLTNLYNSVFTPFVGLFVF- 312
QY 206 IGGAYLLYLNHLGLLLMLHYAVELLSVCSLLYFGDERYOKGLSLWPIVFGSLRLVTLI 265
Db 313 ---FWIYLRHVNVIRILWSVLTFRHGNVYLNFAQOQYKCWISL-PIVVF-----LI 361
QY 266 VSVVTVGLH 274
Db 362 AALQLVNLY 370

RESULT 2

US-08-902-853-7
; Sequence 7, Application US/08902853
; Patent No. 5945330
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: HUMAN LONGEVITY-ASSURANCE PROTEIN HOMOLOGS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902,853
; FILING DATE: Herewith
; CLASSIFICATION: ?
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0345 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 541568
; US-08-902-853-7

Query Match 7.8%; Score 148.5; DB 2; Length 411;
Best Local Similarity 22.0%; Pred. No. 1.1e-08;
Matches 68; Conservative 46; Mismatches 106; Indels 89; Gaps 11;
QY 26 MVSQGVMEFVLGMPFEGTAEMSIVFLTLQHGWWVPAGLPSGSRSTLYHYGVKDLATVFFY 85
Db 91 LVCVTSAYFLSGNRTESNPLHMFVAISYQ-----VDGTDS-----YAKGKDLVSFVFY 139
QY 86 MLVAIIIIHATIQEYVLDKLSRRLOITKQKNEAGOLSVFYI-VSGIWMIL-ASEN 143
Db 140 MIFTFLEFLMDVVIRPFTYVLTNSHQRKMLEQMAIYFCVSPFGLYIMYHSDL 199
QY 144 CLSDPTLLWKSOPHNMFTQMKFFYISQAYW-----FHSPE 181
Db 200 WLFKTKPMYRTPVTNPFLEKIFVLGQAAEWAQACVVLQLEKPRKDYKELVFHHVT 259
QY 182 L-----YFQVKRQDIPGQIYI-----GLHLFH 205
Db 260 LLLIWSYVVFHFTKM-----GLAIYITMDVSDFFLSKLTNLYNSVFTPFVGLFVF- 312
QY 206 IGGAYLLYLNHLGLLLMLHYAVELLSVCSLLYFGDERYOKGLSLWPIVFGSLRLVTLI 265
Db 313 ---FWIYLRHVNVIRILWSVLTFRHGNVYLNFAQOQYKCWISL-PIVVF-----LI 361
QY 266 VSVVTVGLH 274
Db 362 AALQLVNLY 370

RESULT 3

PCT-US95-06725-2
; Sequence 2, Application PC/TUS9506725
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: LAG1: A GENE FOR INCREASING THE
; TITLE OF INVENTION: LONGEVITY OF EUKARYOTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06725
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/253,875 & 08/336,031
; FILING DATE: 03-JUN-1994 & 08-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 93032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-06725-2

Db 271 VFFSCVISLVYHGLMQWVILKIAWLMQVTGTATETLSVAGNIFVSQTEAP-----L 324
QY 140 ASENCLSDPTLLWKSPHNMNT-----FQMKFFYIS-----QLAY 174
Db 325 LIRPYLADMTL---SEVHVMTGGYATIAGSLGAYISFGIDATSLIAASVMAAPCALAL 381
QY 175 WFHSPPELYFQKVRKODIPGQLIYIGLHLFHIGGAYLLYLNLHGLLL---LMLHYAVELL 231
Db 382 SKLVYPEVESFRREE-----GVKLTYGDAQNLEIAASTGAALSVKVVANIAANLI 433
QY 232 SVV-----CSLLYFGDERYQKGLS-----LWPIVIFISG 259
Db 434 AFLAVLDFINAALSWLGDMDVIOGLSFQLCSYILRPVAFMLG 476
RESULT 8
US-08-800-291B-6
; Sequence 6, Application US/08800291B
; Patent No. 6153740
; GENERAL INFORMATION:
; APPLICANT: J.D. Young & C.E. Cass
; TITLE OF INVENTION: CDNA ENCODING NUCLEOSIDE TRANSPORTER
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,291B
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/499,314
; FILING DATE: 7-JULY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07254/044W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 649 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-800-291B-6

Query Match 4.7%; Score 89; DB 4; Length 649;
Best Local Similarity 21.9%; Pred. No. 0.2;
Matches 62; Conservative 49; Mismatches 102; Indels 70; Gaps 11;
QY 30 VGMFFVLGMFEGTAEMSIVFTLQHVVVVPAEGLPSGRTLYHYG-VKDL-----AT 81
Db 211 LGLOFVLGLLVITRTEPGTAFEGWLGQIRIFLSYKAGSSFVGEALVKDVFQVLPPII 270
QY 82 VFYFMLVAIIHATIOEYVLDKLSRRLQTLTKG--KQNKLEAGQLSVFIVSGWMIIL 139
Db 271 VFFSCVISLVYHGLMQWVILKIAWLMQVTGTATETLSVAGNIFVSQTEAP-----L 324
QY 140 ASENCLSDPTLLWKSPHNMNT-----FQMKFFYIS-----QLAY 174

Db 325 LIRPYLADMTL---SEVHVMTGGYATIAGSLGAYISFGIDATSLIAASVMAAPCALAL 381
QY 175 WFHSPPELYFQKVRKODIPGQLIYIGLHLFHIGGAYLLYLNLHGLLL---LMLHYAVELL 231
Db 382 SKLVYPEVESFRREE-----GVKLTYGDAQNLEIAASTGAALSVKVVANIAANLI 433
QY 232 SVV-----CSLLYFGDERYQKGLS-----LWPIVIFISG 259
Db 434 AFLAVLDFINAALSWLGDMDVIOGLSFQLCSYILRPVAFMLG 476
RESULT 9
US-08-800-291B-4
; Sequence 4, Application US/08800291B
; Patent No. 6153740
; GENERAL INFORMATION:
; APPLICANT: J.D. Young & C.E. Cass
; TITLE OF INVENTION: CDNA ENCODING NUCLEOSIDE TRANSPORTER
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,291B
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/499,314
; FILING DATE: 7-JULY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07254/044W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-800-291B-4

Query Match 4.7%; Score 89; DB 4; Length 650;
Best Local Similarity 21.9%; Pred. No. 0.2;
Matches 62; Conservative 49; Mismatches 102; Indels 70; Gaps 11;

QY 30 VGMFFVLGMFEGTAEMSIVFTLQHVVVVPAEGLPSGRTLYHYG-VKDL-----AT 81
Db 212 LGLOFVLGLLVITRTEPGTAFEGWLGQIRIFLSYKAGSSFVGEALVKDVFQVLPPII 271
QY 82 VFYFMLVAIIHATIOEYVLDKLSRRLQTLTKG--KQNKLEAGQLSVFIVSGWMIIL 139
Db 272 VFFSCVISLVYHGLMQWVILKIAWLMQVTGTATETLSVAGNIFVSQTEAP-----L 325
QY 140 ASENCLSDPTLLWKSPHNMNT-----FQMKFFYIS-----QLAY 174
Db 326 LIRPYLADMTL---SEVHVMTGGYATIAGSLGAYISFGIDATSLIAASVMAAPCALAL 382
QY 175 WFHSPPELYFQKVRKODIPGQLIYIGLHLFHIGGAYLLYLNLHGLLL---LMLHYAVELL 231
Db 383 SKLVYPEVESFRREE-----GVKLTYGDAQNLEIAASTGAALSVKVVANIAANLI 434

Qy	232	SSV-----	CSLLYFGDERYQKGLS-----	LWPIVFIG	259
		:	:	:	
		:	:	:	
Db	435	AFLAVLDFINAALSWLGDMDVIOGLSFOLICSYILRPVAFLMG			477

```

RESULT 10
US-09-256-703-2
; Sequence 2, Application US/09256703
; Patent No. 6294379
; GENERAL INFORMATION:
; APPLICANT: Dong, Jian-yun
; APPLICANT: Kan, Yuet Wai
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Efficient AAV Vectors
; FILE REFERENCE: 023070-0849100S
; CURRENT APPLICATION NUMBER: US/09/256,703
; PRIOR FILING DATE: 1999-02-24
; PRIOR APPLICATION NUMBER: US 60/075,980
; PRIOR FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1476
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: truncated cystic fibrosis transmembr
US-09-256-703-2

```

Query Match	4.6%	Score 87;	DB 4;
Best Local Similarity	20.8%	Pred. No. 1.2;	Length 1476;
Matches	83.	Conc.	

[illegible]

```

RESULT 11
US-08-951-912-4
; Sequence 4, Application US/08951912
; Patent No. 5972995
; GENERAL INFORMATION:
; APPLICANT: Fischer, Horst
; APPLICANT: Illek, Beate
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYSTIC
; TITLE OF INVENTION: FIBROSIS THERAPY
; NUMBER OF SEQUENCES: 6

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY, LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/951,912
FILING DATE: 16-OCT-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Mak1, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 200116,403
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1479 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-951-912-4

```

```

Query Match          4.6%;   Score 87;   DB 2;   Length 1479;
Best local Similarity 20.8%;   Pred. No. 1.3;
Matches 83;   Conservative 63;   Mismatches 112;   Indels 142;   Gaps 24;

QY 24 ADMVSCVGMFFVLGMFEGTAEMSIVFLTLOHVVVVPAAEGLPSGSRITLYHYGV-----KD 78
Db 922 ADYLLAMGFGRGLPLVHTLITVSKILHKLHLSVL--QAPNSTLNTLKAGGILNRFRSD 978
QY 79 LA-----TVF-FYMLVAIIH-----ATIQEVLDK-----LSRRLQLTGKQ 115
Db 979 IAILDILLPLTIDFQLLIIVIGAIVAVVLQPIFVATVPVIVATLRLAYFLQTSQ 1038
QY 116 NK-LNEAGOLSWP-YIVSICWMIILASENCISDPTLLWKSQPHMMTFOKMFVISOL- 172
Db 1039 LKQLESGRSPITFHLVLSLGLMTLRA-----FGRQPY-----FETLFHKAALNH 1084
QY 173 -AYWHSFPPELYFORVRQDIPGOLIYIGLHLFHTGGAYLLYL-----NHLGLLLLM-- 223
Db 1085 TANWF-----LYLSTLRWFQMRIFV---IFFIATVFISILTGTGEGRVGIIITLAM 1136
QY 224 -----LHYAV-----ELSSVCSLYFGD-----ERYQKGL-----249
Db 1137 NIMSTLOWAVNNSIDVSLMRKSRVRFKFDIMTEGKPTKSTKPYKNGQLSKYVNIENSH 1196
QY 250 -----SLMPFVIFSGRLVTL--IVSVTVVGLHLAGTNRGNAL-----285
Db 1197 VKDDDIWP-----SGQMTVKDLTAKYTEG-----GNAILNIFSIISPGQRVGLLG 1243
QY 286 ---SGNVNVLAAKIAVLSSSCSIQV-YITWTLTVVWLQRW 321
Db 1244 RTGSGKSTLSAFLRLNLTGEIDGVSW--DSIFLOW 1281

```

RESULT 12
US-09-174-077-4
; Sequence 4, Application US/09174077
; Patent No. 6329422
; GENERAL INFORMATION:
; APPLICANT: Fischer, Horst
; APPLICANT: Illek, Beate
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYSTIC FIBROSIS THERAPY

```

; FILE REFERENCE: 200116.403C1
; CURRENT APPLICATION NUMBER: US/09/174,077
; CURRENT FILING DATE: 1998-10-16
; EARLIER APPLICATION NUMBER: US 08/951,912
; EARLIER FILING DATE: 1997-10-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1479
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-174-077-4

```

Query Match 4.63; Score 87; DB 4; Length 1479;
Best Local Similarity 20.8%; Pred. No. 1.3;
Matches 83; Conservative 63; Mismatches 112; Indels 142; Gaps 24;

QY	24	ADWVCVGHFFVLGLMFECTAEMSIVFLQHGVVVVPAGLPSGSRTRYLYHGV-----KD	78
Dd	922	ADTLAMGEFRLPLVHTLTIVTSKILHHKMLHSVL---QAPMSTLNTLKAGILNRFSSQ	978
QY	79	LA-----TVF-FYMLVAIIIH-----ATIQEYVLDK-----LSRROLTKGKQ	115
Dd	979	IAILDILLPLTIFDFQLGLLVIGAAVAVLQPFIIVATVPVIVAFIMLAYLOTSQQ	1038
QY	116	NK-LNEAGOLSVF-YIVSGIWGMILASBNCILSDPTLLWKSQPHNMFTQMFKFYISOL-	172
Dd	1039	LKOLESEGRSPITHLVTSLGWLTLRA-----FGROPY---FEFLFHKALNLH	1084
QY	173	-AYWFHSFPFLYFOVKVRKODIPCOLIYYIGLHLFHTGGAYLIYL-----NHGLLLLM--	223
Dd	1085	TANWF-----LYLSTLRWPQMIRMEFV---IFFIAVTFSILTAGEGRVGIIITLAM	1136
QY	224	----LHYAV-----ELLSSVCSLLFGD-----ERXQKGL-----	249
Dd	1137	NIMSTLOWAVNSIDVDSLMRSVSRVFIDMPTBGKTPKTCKPYKNQLSKVMIENSH	1196
QY	250	----SLWPIVFISGRIVLTL-IVSVVTVTGLHLAGTNRRGNAL-----	285
Dd	1197	VKKDDIWP---SGQOMTKDLTAKYTEG-----GNILENISFISPGORVGLLG	1243
QY	286	----SGNVNVLAKIAVLSSCSCIQV-YITWTLTVTVMLOWR	321
Dd	1244	RTSGKSTLSLAFLRLNTEGIIQIDGYSW--DSTITLOOW	1281

```

RESULT 13
US-07-637-621-2
; Sequence 2, Application US/07637621
; Patent No. 5407796
; GENERAL INFORMATION:
; APPLICANT: cutting, gary
; APPLICANT: antonarakis, stylianos e
; APPLICANT: kazazian jr., haig h
; TITLE OF INVENTION: CYSTIC FIBROSIS MUTATION CLUSTER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie and Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07637,621
; FILING DATE: 19910104
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

```

NAME: kagan, sarah a
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.030010
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9100
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1480 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: HOMO SAPIENS
US-07-637-621-2

```

Query Match 4.6%; Score 87; DB 1; Length 1480;
Best Local Similarity 20.8%; Pred. NO. 1.3;

```

Qy   24 ADMVSCVMFEVLGLMEPEGTAEMSVIFLTLQHGVVVPAEGLPGSGRPLYHYGV-----KD 78
      || : || | : || : || : || : || : || : || : || : || : || : || : ||
Db   923 ADTLLAMGFFRGFLPLVFTLTVSKILHKKMLHSVL--QAPMSTLNTLKAGGILLNRESKD 979
      || : || | : || : || : || : || : || : || : || : || : || : ||
Qy   79 LA-----TVF-FYMLVAIIIH-----ATIOBYLDK-----LSRRQLTGKGK 115
      || : || | : || | : || | : || | : || | : || | : || | : || | : ||
Db   980 IAILDDLLPTIFIDLIOLLIVIGAIWAVLQPIFVATVPVIAMIMRAYFLTQSQQ 1039
      || : || | : || | : || | : || | : || | : || | : || | : || | : ||
Qy   116 NK-LNEAQLSVF-YIYSWGWMILIASECLSDPTLLQKSQNMNMTFOMKFFYSIQ- 172
      || : || | : || : || : || : || : || : || : || : || : || : ||
Db   1040 LKQLESEGRSPITHLVTSKLWLTRA-----FGROPY-----FETLFHKALNLH 1085
      || : || | : || : || : || : || : || : || : || : || : || : ||
Qy   173 -AYWFHSPELYFOKKRKQDIPGOLIVIGHLPHIGAYLLYL-----NHLGLLLML-- 223
      || || || : || : || : || : || : || : || : || : || : || : ||
Db   1086 TANFW----LYLSTLRWFQMRIEMIEV---IFFIAVTFSITLTGTGEGRVGILTLM 1137
      || : || | : || : || : || : || : || : || : || : || : || : ||
Qy   224 ----LHYAV-----ELLSSVCSLLYFGD-----ERYQKGL----- 249
      || : || | : || : || : || : || : || : || : || : || : || : ||
Db   1138 NIMSTLOWANVSITVDLSMRSSVRPFKFIDMPTEGPKTKPYRNKGQLSKVMIIENSH 1197
      || : || | : || : || : || : || : || : || : || : || : || : ||
Qy   250 ----SLAPIVFISGRVLTL--IVSWVVVGLHLACTNRNGNAL----- 285
      || || || : || : || : || : || : || : || : || : || : || : ||
Db   1198 VKRDDIWP----SGQMTVKRDLPKYTEG-----GNAILNISFSIPGORVCLLG 1244
      || || || : || : || : || : || : || : || : || : || : || : ||
Qy   286 ---SGNVNVLAAKIATVLUSSCSIOQV-YITWTLTFTVVLQRW 321
      || : || | : || : || : || : || : || : || : || : || : || : ||
Db   1245 RTSCKGKTLLSAFULRLNTEGETOIDCVSNW--DSITLQOOW 1282

```

```

RESULT 14
US-08-136-742A-2
; Sequence 2, Application US/08136742A
; Patent No. 5670488
; GENERAL INFORMATION:
; APPLICANT: Gregory, R.J., Armentano, D., Couture, L.,
; APPLICANT: A.E.
; TITLE OF INVENTION: GENE THERAPY FOR CYSTIC FIBROSIS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BRUMBAUGH, GRAVES, DONOHUE & RAYMOND
; STREET: 30 ROCKEFELLER PLAZA
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/136.742A

```

FILING DATE: 02-DEC-1993
 CLASSIFICATION: 514
 PRIOR APPLICATION NUMBER: 514
 APPLICATION DATA:
 FILING DATE: 02-DEC-1992
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Seide, Rochelle K.
 REGISTRATION NUMBER: 32,300
 REFERENCE/DOCKET NUMBER: A30668 (Genzyme Dkt. IG4-9.11)
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 408-2500
 TELEFAX: (212) 763-2519
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1480 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-136-742A-2

Query Match
 Best Local Similarity 4.6%; Score 87; DB 1; Length 1480;
 Matches 83; Conservative 63; Mismatches 112; Indels 142; Gaps 24;
 QY 24 ADMVSCVGMFFVLGLMFEGETAEMSIIVFLTQHGCVVPAEGLPSGSRITLYHYGV-----KD 78
 Db 923 ADTLAMGFFRGLPLVHTLTVSKILHMLHSVL---QAPMSTLNTLKAGGILNRFSD 979
 QY 79 LA-----TVP-FYMLVAIIH-----ATIQEYVLDK-----LSRRLQITKQK 115
 Db 980 IAILDDLLPTIFDFIQLLLIVIGATAVAVLQPIYFVATVPVIVAFIMLRAFLQTSQ 1039
 QY 116 NK-LNEAGQLSVF-YIVSGIWMGMIILASENCSDPTLLWKSOPHNMFTQMKFFYISOL- 172
 Db 1040 LKOLESEGRSPIFTHLVTSKGLWTLRA-----FGRQPY-----FETLFHKAALNH 1085
 QY 173 -AYWFHSEPELYFQVKRQDIPQOLIYIGLHFLHIGGAYLLYL-----NHLGILLM-- 223
 Db 1086 TANWF-----LYLSTURWFQMRIMFV---IFFIAVTFISLTGTGEGRVIILTLAM 1137
 QY 224 -----LHYAV-----ELLSSVCSLLYFGD-----ERYQGL----- 249
 Db 1138 NIMSTLQWAVNSSIDVSLMRSVSRVFKFIDMPTGKPTKSTPKYKNGOLSKVMIENSH 1197
 QY 250 ----SLWPVIFSGRLVTL--IVSVTVVGLHLAGTNRGNAL----- 285
 Db 1198 VKKDDIWP-----SGQMTVKDLTKAYTEG-----GNAILNIFSISPCQRVGLLG 1244
 QY 286 ---SGNVNVLAAKIAVLSSSCSIQV-YITWTLTVMWLQW 321
 Db 1245 RTGSGKSTLLSAFLRLNTEGEIQIDGVSW--DSITLQW 1282

RESULT 15
 US-08-135-809A-2
 Sequence 2, Application US/08135809A
 Patent No. 5688677
 GENERAL INFORMATION:
 APPLICANT: CHENG, SENG H.
 APPLICANT: DITULLIO, PAUL
 APPLICANT: EBERT, KARL M.
 APPLICANT: MEADE, HARRY M.
 APPLICANT: SMITH, ALAN E.
 TITLE OF INVENTION: DROXYRIBONUCLEIC ACIDS CONTAINING
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENZYME CORPORATION
 STREET: ONE MOUNTAIN ROAD
 CITY: FRAMINGHAM
 STATE: MASSACHUSETTS

COUNTRY: USA
 ZIP: 01701
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/135,809A
 FILING DATE: 13-OCT-1993
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: LASSEN, ELIZABETH
 REGISTRATION NUMBER: 31,845
 REFERENCE/DOCKET NUMBER: IG4-9.12
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (508) 872-8400
 TELEFAX: (508) 872-5415
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1480 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-135-809A-2

Query Match
 Best Local Similarity 4.6%; Score 87; DB 1; Length 1480;
 Matches 83; Conservative 63; Mismatches 112; Indels 142; Gaps 24;
 QY 24 ADMVSCVGMFFVLGLMFEGETAEMSIIVFLTQHGCVVPAEGLPSGSRITLYHYGV-----KD 78
 Db 923 ADTLAMGFFRGLPLVHTLTVSKILHMLHSVL---QAPMSTLNTLKAGGILNRFSD 979
 QY 79 LA-----TVP-FYMLVAIIH-----ATIQEYVLDK-----LSRRLQITKQK 115
 Db 980 IAILDDLLPTIFDFIQLLLIVIGATAVAVLQPIYFVATVPVIVAFIMLRAFLQTSQ 1039
 QY 116 NK-LNEAGQLSVF-YIVSGIWMGMIILASENCSDPTLLWKSOPHNMFTQMKFFYISOL- 172
 Db 1040 LKOLESEGRSPIFTHLVTSKGLWTLRA-----FGRQPY-----FETLFHKAALNH 1085
 QY 173 -AYWFHSEPELYFQVKRQDIPQOLIYIGLHFLHIGGAYLLYL-----NHLGILLM-- 223
 Db 1086 TANWF-----LYLSTURWFQMRIMFV---IFFIAVTFISLTGTGEGRVIILTLAM 1137
 QY 224 -----LHYAV-----ELLSSVCSLLYFGD-----ERYQGL----- 249
 Db 1138 NIMSTLQWAVNSSIDVSLMRSVSRVFKFIDMPTGKPTKSTPKYKNGOLSKVMIENSH 1197
 QY 250 ----SLWPVIFSGRLVTL--IVSVTVVGLHLAGTNRGNAL----- 285
 Db 1198 VKKDDIWP-----SGQMTVKDLTKAYTEG-----GNAILNIFSISPCQRVGLLG 1244
 QY 286 ---SGNVNVLAAKIAVLSSSCSIQV-YITWTLTVMWLQW 321
 Db 1245 RTGSGKSTLLSAFLRLNTEGEIQIDGVSW--DSITLQW 1282

Search completed: September 6, 2002, 17:10:38
 Job time: 7972 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 15:18:41 ; Search time 52.02 seconds
(without alignments)
670.520 Million cell updates/sec

Title: US-09-807-470-2

Perfect score: 1892

Sequence: 1 MGLRKKNARNPPVLSHEFMV.....NGVENPNRIDSPPKKKKAP 363

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1150.5	60.8	374	2 S21736	translocating chai
2	1134.5	60.0	374	2 S30034	translocating chai
3	441.5	23.3	371	2 T19417	hypothetical prote
4	434.5	23.0	373	2 T19419	hypothetical prote
5	148.5	7.8	411	2 S46800	LAG1 protein - yea
6	137	7.2	418	2 S30134	hypothetical prote
7	133	7.0	357	2 T40389	longevity assuranc
8	125	6.6	390	2 T38012	longevity-assuranc
9	116.5	6.2	696	2 B86726	hypothetical prote
10	107.5	5.7	614	2 A98845	Na+/H+ antiporter
11	103.5	5.5	393	2 E96954	Na/H antiporter (n
12	102.5	5.4	522	2 B83987	ABC transporter (p
13	101	5.3	308	2 H86268	hypothetical prote
14	100.5	5.3	360	2 T27324	hypothetical prote
15	100.5	5.3	397	2 D71467	probable tyrosine
16	99.5	5.3	395	2 D81729	Mtr/TnaB/Tyros perm
17	99	5.2	286	2 AG0403	anaerobic dimethyl
18	98.5	5.2	722	2 G83685	hypothetical prote
19	97.5	5.2	397	2 T00098	hypothetical prote
20	97	5.1	372	2 AH0703	probable membrane
21	96.5	5.1	534	2 D71698	cytochrome-c oxida
22	95.5	5.0	333	2 A10050	probable ABC trans
23	95.5	5.0	411	2 G90154	arsenite transport
24	95	5.0	370	2 H64926	probable membrane
25	95	5.0	370	2 C90528	hypothetical prote
26	95	5.0	370	2 G85776	hypothetical prote
27	95	5.0	509	2 G81929	probable iron-upta
28	94.5	5.0	540	1 I49454	sterol O-acetyltrans
29	94.5	5.0	547	2 T27253	hypothetical prote

30 94.5 5.0 788 2 S48191 probable ubiquinol
31 94 707 2 T09340 hypothetical prote
32 94 1224 2 H96615 hypothetical prote
33 94 1450 2 JC6139 cystic fibrosis tr
34 93.5 4.9 519 2 D85437 hypothetical prote
35 93.5 4.9 521 2 A32431 cytochrome-c oxida
36 93 4.9 417 2 AB0682 probable membrane
37 93 4.9 495 2 E81251 NADH dehydrogenase
38 93 4.9 515 2 B90504 amino acid transpo
39 93 4.9 531 2 T40575 major facilitator
40 93 4.9 532 1 E69821 multidrug resistanc
41 93 4.9 542 2 E90604 hypothetical prote
42 92.5 4.9 353 2 AF0676 hydrogenase-1 oper
43 92.5 4.9 723 2 T21869 hypothetical prote
44 92 4.9 345 2 AC1252 conserved hypothet
45 92 4.9 447 2 D90012 hypothetical prote

ALIGNMENTS

RESULT 1

S21736

translocating chain-associating membrane protein - dog

C;Species: Canis lupus familiaris (dog)

C;Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 24-Sep-1999

C;Accession: S21736

R;Goerlich, D.; Hartmann, E.; Prehn, S.; Rapoport, T.A.

Nature 357, 47-52, 1992

A;Title: A protein of the endoplasmic reticulum involved early in polypeptide translo

A;Reference number: S21736; MUID:92244357

A;Accession: S21736

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-374 <GOE>

A;Cross-references: EMBL:X63678; NID:g941; PIDN:CAA45217.1; PID:g942

C;Superfamily: translocating chain-associating membrane protein

C;Keywords: transmembrane protein

Query Match 60.8%; Score 1150.5; DB 2; Length 374;
Best Local Similarity 59.8%; Pred. No. 1.8e-92;
Matches 223; Conservative 58; Mismatches 81; Indels 11; Gaps 4;

QY 1 MGLRKKNARNPPVLSHEFMVQNHADIVSCVGMFEFVLGLMFEGETAEMSIIVLTLOHGVVVP 60

Db 1 MAIRKSTKSPVLSHEFILLQNHADIVSCVAVFLLGLMFEITAKASIIIVTLQNTVLP 60

QY 61 A-EGLPSSRTLYHYGVKDLATVFFYMLVAIIIIHATIQEYVLDKLSRLQLTKGKQKLN 119

Db 61 ATEQATSTSLYYGIGKDLATVFFYMLVAIIIIHATIQEYVLDKLSRLQLTKGKQKLN 120

QY 120 EAGLSVFIYVSGTWGMIIASENCLSDPTLLNKSQPINMTFQMKFYISQLAYWFHSF 179

Db 121 ESGLSAFYLFSCIWGTFLISENYISDPTILWRAYPHNLTQMKFFYIAQLAYWFHAF 180

QY 180 PELYFQVRKODIPQLYIGLHLFHIGVALLVNLHGLLLHLHVAVELLSVCSLLY 239

Db 181 PELYFQTKREDIPQLYIGLHLFHIGVALLVNLHGLLLHLHVAVELLHISRLFY 240

QY 240 FGDERYQKGLSLWPIVFTISGRVLTIVSVVTVGLHLA-GTNRNGNALSGNVNLAATIAV 298

Db 241 FSDKYQKGLSLWAVLVGLGRLLTLILSVLVGFLARAENQKLDIFSAGNENLVAVRAV 300

QY 299 LSSSCSIQVYITWTLTTVWLQRLWLEDANLHVCGRKKR-----SRSRKSTENGVE---NP 349

Db 301 LASICITQAFMMWKFNFQLRRRREHSTFQAPVVKKKPTVTKGRSRKGTENGNGVTVS 360

QY 350 NRIDSPPKKEKA 362

Db 361 NGADSPRNREKS 373

RESULT 2

S30034
translocating chain-associating membrane protein - human
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-Sep-1999
C:Accession: S30034
R:Goerlich, D.; Hartmann, E.; Prehn, S.; Rapoport, T.A.
A:Title: A protein of the endoplasmic reticulum involved early in polypeptide translocation
A:Reference number: S21736; MUID:92244357
A:Accession: S30034
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-374 <GOE>
A:Cross-references: EMBL:X63769; NID:g37264; PIDN:CAA45218.1; PID:g37265
C:Superfamily: translocating chain-associating membrane protein

Query Match 60.0%; Score 1134.5; DB 2; Length 374;
Best Local Similarity 58.7%; Pred. No. 4.6e-91;
Matches 219; Conservative 61; Mismatches 82; Indels 11; Gaps 4;
QY 1 MGLRKNARPPVLSHEFMVQNHADMYSCVGMFFVLGLMFEGETAEMSIIVLTLOHGYYVVP 60
Db 1 MAIRKKTSPVLSHEFVLQNHADIVSCVAMVFLGLMFEITAKASIIFFVTLQYVNTLP 60
QY 61 A-EGLPSSGRTLYHY--GVKDLATVFFVLMVAIIHATIQEYVLDKLSRRLOLTGKQNKLN 119
Db 61 ATEQATSESVLYYIGIKDLATVFFVLMVAIIHATIQEYVLDKLSRRLOLTGKQNKFN 120
QY 120 EAGLSVFIYVSGIWMIIASENCLSDPTLLWKSSOPHNMFTQMKFFYISQLAYWFHSF 179
Db 121 ESGLSAFYLFACVWGTFILISENYSIDPTILWRAPYHNLMTQMKFFYISQLAYWLF 180
QY 180 PELYFQVRKQDIPGOLYIIGLHFGGAYLLYLNHLGLLLMLHYAVELLSSVCLLY 239
Db 181 PELYFQTKKEDIPQLVYIGLFLHAGAYLLNHLGLVLLVLYFEFLFHSIRLPY 240
QY 240 FGERYQKGLSLMPTVIFISGRVLTIVSVVTVGLHLA-GTNRGNALSGNVNVAIAKIAV 298
Db 241 FSNERYQKGLSLMPTVIFISGRVLTIVSVVTVGLHLA-GTNRGNALSGNVNVAIAV 300
QY 299 LSSSCSIQVYITWTLTVWLQRLWLEDANLHVCGRKR-----SRSKKTENGVE---NP 349
Db 301 LASICVTQAFMWMKFINFQLRRHSAFAQPAVKKKPTVTKGRSSKKGTVNGVNGTLTS 360
QY 350 NRIDSPPKKKEKA 362
Db 361 NVADSPRNKKEKS 373

RESULT 3

T19417
hypotheical protein C24F3.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T19417
R:McMurray, A.
A:Title: A protein of the endoplasmic reticulum involved early in polypeptide translocation
A:Reference number: S21736; MUID:92244357
A:Accession: T19417
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-371 <WIL>
A:Cross-references: EMBL:AL022716; PIDN:CAA18770.1; GSPDB:GN00022; CESP:C24F3.1a
A:Experimental source: clone C24F3
C:Genetics:
A:Gene: CESP:C24F3.1a
A:Map position: 4
A:Introns: 114/3; 158/2; 363/3
C:Superfamily: translocating chain-associating membrane protein

Query Match 23.3%; Score 441.5; DB 2; Length 371;
Best Local Similarity 30.0%; Pred. No. 9.8e-31;
Matches 112; Conservative 76; Mismatches 152; Indels 33; Gaps 11;
QY 2 GLRKNARPPVLSHEFMVQNHADMYSCVGMFFVLGLMFEGETAEMSIIVLTLOHGYYVVP 61
Db 7 GSKASKKPPVLSHEFVLQNHADIMSCVVMVFIIVGLMFLPLTHLSLSLFTAPQYNGTYV 66
QY 62 EGLPSGRTLYHY--GVKDLATVFFVLMVAIIHATIQEYVLDKLSRRLOLTGKQNKLN 119
Db 67 AVEQGOEREVHGYSGLDLPALFFYSVCWIVVHVVQVEYGLDKISKKTHLSKVSFTKFG 126
QY 120 EAGLSVFIYVSGIWMIIASENCLSDPTLLWKSSOPHNMFTQMKFFYISQLAYW 175
Db 127 ESPH-QMFTVYSIAHAFYIVSERLEDFSEKSVWLGYPTEHVRVMSAAVLYFIQISYW 185
QY 176 FHSFPELYFQVRKQDIPGOLYIIGLHFGGAYLLYLNHLGLLLMLHYAVELLSSVC 235
Db 186 IHQFPFYLQKLRDEIRKQSVQALIHAFISTAYFNFTRVGLALITLEYITQLIFHIA 245
QY 236 SLLYFGDERYQKGLS-----LMPVIFISGRVLTIVSVTV--VGLHLAGTNRGNALSGN 288
Db 246 RFAHFVG---RKGLSDPAFKLENGSEFVLRLGSIILIAVMTFWYGLRQA-ESPVDISAGN 301
QY 289 VNVLAAKIAVLSSCSIQYIITWTLTVWLQRLWLEDANLHVCGRKRKRSGTENGVEN 348
Db 302 FNTAVIRLNVLLAVLLQLFLLYSFVV-----FHM-GRFESNAKKEKKKSA 349
QY 349 PNRIIDSPPKKKEK 361
Db 350 AAHV---PKKEK 359

RESULT 4

T19419
hypotheical protein C24F3.1b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T19419
R:McMurray, A.
A:Title: A protein of the endoplasmic reticulum involved early in polypeptide translocation
A:Reference number: Z19122
A:Accession: T19419
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-373 <WIL>
A:Cross-references: EMBL:AL022716; PIDN:CAA18772.1; GSPDB:GN00022; CESP:C24F3.1b
A:Experimental source: clone C24F3
C:Genetics:
A:Gene: CESP:C24F3.1b
A:Map position: 4
A:Introns: 114/3; 158/2; 365/3
C:Superfamily: translocating chain-associating membrane protein

Query Match 23.0%; Score 434.5; DB 2; Length 373;
Best Local Similarity 30.1%; Pred. No. 4e-30;
Matches 113; Conservative 75; Mismatches 151; Indels 37; Gaps 12;
QY 2 GLRKNARPPVLSHEFMVQNHADMYSCVGMFFVLGLMFEGETAEMSIIVLTLOHGYYVVP 61
Db 7 GSKASKKPPVLSHEFVLQNHADIMSCVVMVFIIVGLMFLPLTHLSLSLFTAPQYNGTYV 66
QY 62 EGLPSGRTLYHY--GVKDLATVFFVLMVAIIHATIQEYVLDKLSRRLOLTGKQNKLN 119
Db 67 AVEQGOEREVHGYSGLDLPALFFYSVCWIVVHVVQVEYGLDKISKKTHLSKVSFTKFG 126
QY 120 EAGLSVFIYVSGIWMIIASENCLSDPTLLWKSSOPHNMFTQMKFFYISQLAYW 172
Db 127 ESPH-QMFTVYSIAHAFYIVSER-LEDFSEKSVFSVWLGYPTEHVRVMSAAVLYFIQI 184
QY 173 AYNFHSPELYFQVRKQDIPGOLYIIGLHFGGAYLLYLNHLGLLLMLHYAVELLSS 232

Db 185 SYWIRHQPEFYLOKLKRDRIKQSVQAIIHLIAFISIAFYFNTRVGLALITLEYITQLIF 244
QY 233 SVCSLLYFGDERYQKGLS-----LWPIVFISGRVLTIVSVVT--VGLHLACTNRNGNAL 285
Db 245 HIRAFHFGV---RGLSDPAKFLNGSFVLVRLGSLIIIVMTFVYGLRQA-ESFPFVDIS 300
QY 286 SGNVNVLAAKIAVLSSSSQIYIITWTITVVLQRLWLEDANLHVCGVGRKRKRSRSGKTENG 345
Db 301 AGNFTAVIRNLVALLAVVLQLFLYSFV-----FHM-GRFPESNAKKEKKKS 348
QY 346 VENPNRIDSPPKKEK 361
Db 349 AAAAAAV---PKKERK 361

RESULT 5

S46800
LAG1 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YHL003c
C:Species: Saccharomyces cerevisiae
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 23-Mar-2001
C:Accession: S46800; A54012
R:Favetto, T.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of S. cerevisiae cosmid 9780.
A:Reference number: S46797
A:Accession: S46800
A:Molecule type: DNA
A:Residues: 1-411 <FAV>
A:Cross-references: EMBL:U10555; NID:G500813; PIDN:AAB68429.1; PID:G500820; MIPS:YHL003c
R:D'mello, N.P.; Childress, A.M.; Franklin, D.S.; Kale, S.P.; Pinswasdi, C.; Jazwinski, J. Biol. Chem. 269, 15451-15459, 1994
A:Title: Cloning and characterization of LAG1, a longevity-assurance gene in yeast.
A:Reference number: A54012; MUID:94253121
A:Accession: A54012
A:Molecule type: DNA
A:Residues: 1-172, 'IV', 175-219, 'C', 221-300, 'TEISGI', 314, 'EKQE', 315, 'DSNDNPTE', 324, 'A', 325, 'A', 381, 'AGOR', 386, 'L', 388, 'NRLARNEK' <DAM>
A:Cross-references: GB:U08133
C:Genetics:
A:Gene: SGD:LAG1
A:Cross-references: SGD:S0000995; MIPS:YHL003c
A:Map position: 8L
C:Function:
A:Description: involved in dertermination of longevity
C:Superfamily: hypothetical protein YKL008c
C:Keywords: transmembrane protein

Query Match 7.8%; Score 148.5; DB 2; Length 411;
Best Local Similarity 22.0%; Pred. No. 3.6e-05;
Matches 68; Conservative 46; Mismatches 106; Indels 89; Gaps 11;
QY 26 MYSCVGMFFVLGLMFEPTGAEMSIIVTLTQHGVVVPAEGLPSGSRITLYHYGVKDLATVFY 85
Db 91 LVCVVSAYFLSGNRTESNPLHMFVAISYQ-----VDGTDSD-----YAKGKDLSEVFY 139
QY 86 MLVAIIIRATQEVYLDKLSRLRLQTKQKQNLNAGQSLSVFYI-VSGTWGMIL-ASEN 143
Db 140 MIFFTFLREFLMDVIRPFTVNLVTSBHRQRMLEQMYAIFCVGSGPFGLYIMYHSDL 199
QY 144 CLSDPTLLWKSOPHNMTFMQFFYISQLAYW-----FHSFPE 181
Db 200 WLFTKPMYRTPVTNTNPLFKIFILGQAAFAQAQCVLVQLKPRKDYKELVFHHIVT 259
QY 182 L-----YFQVKRKQDIPGQIYI-----GLHLFPH 205
Db 260 LLLIWSYVFHETKM-----GLAIYIMDVSDFFLSLKTLYNLNSVTPPFVGLFVP- 312
QY 206 IGGAYLLYNHLGLLLMLHYAVELLSSVSLLYFGDERYQKGLSLWPIVFISGRVLTLI 265
Db 313 ----FWIVLRHVNVIRILWSVLTEFRHGNVYLNATQYKQKWISL-PIVFV-----LI 361

QY 266 VSVVTVGLH 274
Db 362 AALQLVNLY 370
RESULT 6
S30134
hypothetical protein YKL008c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YKL156
C:Species: Saccharomyces cerevisiae
C:Date: 18-Jun-1993 #sequence_revision 18-Jun-1993 #text_change 24-Sep-1999
C:Accession: S30134; S37819; S17017
R:Boyer, J.; Pascolo, S.; Richard, G.F.; Dujon, B.
Yeast 9, 279-287, 1993
A:Title: Sequence of a 7.8 kb segment on the left arm of yeast chromosome XI reveals
mammalian UOG-1 gene.
A:Reference number: S30132; MUID:93255906
A:Accession: S30134
A:Molecule type: DNA
A:Residues: 1-418 <BOY>
A:Cross-references: GB:S59773; NID:G300231; PIDN:AAC60549.1; PID:G300232
A:Experimental source: strain S288C
R:Boyer, J.; Pascolo, S.; Richard, G.F.; Ghazvini, M.; Collea, L.; Thierry, A.; Men
submitted to the Protein Sequence Database, March 1994
A:Reference number: S37813
A:Accession: S37819
A:Molecule type: DNA
A:Residues: 1-418 <BO2>
A:Cross-references: EMBL:Z28008; NID:G485982; PIDN:CAA81843.1; PID:G485983; MIPS:YKL0
A:Experimental source: strain S288C
R:Amatruda, J.F.; Gattermeier, D.G.; Cooper, J.A.
submitted to the EMBL Data Library, August 1991
A:Description: Yeast capping protein.
A:Reference number: S17016
A:Accession: S17017
A:Molecule type: DNA
A:Residues: 1-149 <AMA>
A:Cross-references: EMBL:X61398; NID:G455515; PIDN:CAA43670.1; PID:G3445
C:Genetics:
A:Map position: 11L
A:Superfamily: hypothetical protein YKL008c
C:Keywords: transmembrane protein

Query Match 7.2%; Score 137; DB 2; Length 418;
Best Local Similarity 23.5%; Pred. No. 0.00037;
Matches 61; Conservative 44; Mismatches 93; Indels 62; Gaps 11;
QY 67 GSRTLYHYGVKDLATVFYVYMLVAIIIHATIQEVYLDKLSRRLQTLTKGQKQNLNAGQSLV 126
Db 121 GDTNAYKGINDLCFVYFYIMIFFTLREFLMDVIRPFAIRLHVTSKHKRIKRIEOMYAI 180
QY 127 FYI-VSGIWMIL-ASENCLSDPTLLWKSOPHNMTFMQFFYISQLAYW----- 175
Db 181 FYTGVSGPGIGCMYHSDLWFFNTKAMRYTDPFTNPFLFKVYFGAFAQAQACILVL 240
QY 176 -----FHSFPEL-----YFQVKRKQDIPGQIYIYGLHFLHIGGAYLLYNH 216
Db 241 QLEKPRKDHNEITFHIVTLLLIWSYVFHETKMGLP---IYITMDVSDFLSFSKTLNY 297
QY 217 L-GLL-----LLMLHYA-VELLSVCS-----LLYFGDERYQKGLSWPI 254
Db 298 LDSGLAFFSFAIVVAMIYLRHYINLILWSVLTPQRTGNYVNLNATQYKQKWISL-PI 356
QY 255 VFISGRVLTIVSVTVGLH 274
Db 357 VFV-----LIGALQLVNLY 370

RESULT 7
T40389
longevity assurance protein homolog SPBC3E7.15c - fission yeast (Schizosaccharomyces
C:Species: Schizosaccharomyces pombe

[illegible]

Experimental source: CROSTRIUM acetylcholine A100024

Genetics:

444

A; Experiment
C; Genetics

190 TYR

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

11-215

Search completed: September 6, 2002, 17:11:51
Job time: 6790 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 17:10:42 ; Search time 26.42 Seconds
(without alignments)
531.991 Million cell updates/sec

Title: US-09-807-470-2

Perfect score: 1892

Sequence: 1 MGLRKNRNPNVLSHEFMV.....NGVENPNRIDSPKKKEKAP 363

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1145.5	60.5	373	1	TRAM_CANFA
2	1129.5	59.7	373	1	TRAM_HUMAN
3	1084.5	57.3	358	1	TRAM_BOVIN
4	754	39.9	370	1	Y557_HUMAN
5	148.5	7.8	411	1	LAG1_YEAST
6	137	7.2	418	1	YK48_YEAST
7	133	7.0	384	1	YHXP_SCHPO
8	125	6.6	390	1	LAG1_SCHPO
9	97.5	5.2	533	1	GLPT_HUMAN
10	97	5.1	1450	1	CFTR_RABIT
11	96.5	5.1	534	1	COX1_RICPR
12	95	5.0	370	1	YDIK_ECOLI
13	94.5	5.0	540	1	SOAL_MOUSE
14	94.5	5.0	788	1	QOXM_SULAC
15	94	5.0	272	1	ATP6_BUCAP
16	93.5	4.9	521	1	COX1_APILI
17	93	4.9	532	1	YHCA_BACSU
18	91.5	4.8	369	1	Y316_MYCPN
19	91	4.8	307	1	OC22_HUMAN
20	91	4.8	624	1	NKX3_RAT
21	91	4.8	645	1	NKX3_MOUSE
22	90	4.8	513	1	COX1_RABIT
23	90	4.8	514	1	COX1_CERSI
24	90	4.8	514	1	COX1_EQUAS
25	90	4.8	514	1	COX1_HORSE
26	90	4.8	514	1	COX1_RHITU
27	90	4.8	516	1	COX1_GADMO
28	89	4.7	360	1	MRAY_PASMU
29	89	4.7	430	1	SECY_BACHD
30	89	4.7	499	1	MYIN_AQUAE
31	89	4.7	514	1	COX1_BOVIN
32	89	4.7	514	1	COX1_CANFA
33	89	4.7	514	1	COX1_FELCA

34	89	4.7	514	1	COX1_SHEEP	078749	ovis aries
35	89	4.7	644	1	NKX3_HUMAN	09hc58	homo sapien
36	89	4.7	649	1	CNTL_HUMAN	000337	homo sapien
37	88.5	4.7	1481	1	CFTR_BOVIN	P33071	bos taurus
38	88.5	4.7	1481	1	CFTR_SHEEP	Q00555	ovis aries
39	88	4.7	433	1	SECY_RICPR	Q9zcs5	rickettsia
40	88	4.7	515	1	COX1_HIPAM	Q9zzy9	hippopotamu
41	88	4.7	531	1	YDFG_SCHPO	010487	schizosacch
42	87.5	4.6	297	1	CYAA_BACFI	Q04443	bacillus fi
43	87.5	4.6	515	1	COX1_CHICK	P18943	gallus gall
44	87	4.6	432	1	REBX_SALTI	Q99191	salmoneilla
45	87	4.6	514	1	COX1_PIG	079876	sus scrofa

ALIGNMENTS

```
RESULT 1
TRAM_CANFA
ID TRAM_CANFA STANDARD; PRT; 373 AA.
AC Q01685;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE TRAM protein (Translocating chain-associating membrane protein).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-26 AND 164-184.
RC TISSUE=Kidney;
RX MEDLINE=92244357; PubMed=1315422;
RA Goerlich B., Hartmann E., Frenn S., Rapoport T.A.;
RT "A protein of the endoplasmic reticulum involved early in polypeptide
RT translocation.";
RL Nature 357:47-52(1992).
CC -!- FUNCTION: STIMULATORY OR REQUIRED FOR THE TRANSLOCATION OF
CC SECRETORY PROTEINS ACROSS THE ER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum membrane.
CC -!- SIMILARITY: BELONGS TO THE LASSI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X63678; CAA45217.1; -.
DR PIR: S21736; S21736.
DR Endoplasmic reticulum; Transmembrane; Glycoprotein; Translocation.
FT INIT_MET 0
FT DOMAIN 1 28 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 29 49 POTENTIAL.
FT DOMAIN 50 75 LUMENAL (POTENTIAL).
FT TRANSMEM 76 96 POTENTIAL.
FT DOMAIN 97 120 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 121 141 POTENTIAL.
FT DOMAIN 142 158 LUMENAL (POTENTIAL).
FT TRANSMEM 159 179 POTENTIAL.
FT DOMAIN 180 191 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 192 212 POTENTIAL.
FT DOMAIN 213 216 LUMENAL (POTENTIAL).
FT TRANSMEM 217 237 POTENTIAL.
FT DOMAIN 238 250 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 251 271 POTENTIAL.
FT DOMAIN 272 296 LUMENAL (POTENTIAL).
FT TRANSMEM 297 317 POTENTIAL.
FT DOMAIN 318 373 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 55 N-LINKED (GLCNAC... ) (PROBABLE).
SQ SEQUENCE 373 AA; 43029 MW; 1D85808E1D80E835 CRC64;
```



```

NCBI_TaxID=9913;
SEQUENCE FROM N.A.
Clark T.G., Morris J., Akamatsu M., McGraw R.A., Ivarie R.D.;
"Cloning and sequence analysis of a bovine tram cDNA.";
Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: STIMULATORY OR REQUIRED FOR THE TRANSLLOCATION OF
SECRETORY PROTEINS ACROSS THE ER MEMBRANE (By similarity).
-!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM MEMBRANE (By
similarity).
-!- SIMILARITY: BELONGS TO THE LASS1 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U19578; AAG10391.1; -
Endoplasmic reticulum; Transmembrane; Glycoprotein; Translocation.
KW NON_TER
FT DOMAIN <1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 POTENTIAL.
FT DOMAIN 35 60 LUMENAL (POTENTIAL).
FT TRANSMEM 61 81 POTENTIAL.
FT DOMAIN 82 105 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 106 126 POTENTIAL.
FT DOMAIN 127 143 LUMENAL (POTENTIAL).
FT TRANSMEM 144 164 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 165 176 POTENTIAL.
FT TRANSMEM 177 197 LUMENAL (POTENTIAL).
FT DOMAIN 198 201 POTENTIAL.
FT TRANSMEM 202 222 POTENTIAL.
FT DOMAIN 223 235 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 236 256 POTENTIAL.
FT DOMAIN 257 281 LUMENAL (POTENTIAL).
FT TRANSMEM 282 302 POTENTIAL.
FT DOMAIN 303 358 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 40 40 N-LINKED (GLCNAC... ) (PROBABLE).
SQ SEQUENCE 358 AA; 41403 MW; 25CF9930C4CDDA15 CRC64;

Query Match 57.3%; Score 1084.5; DB 1; Length 358;
Best Local Similarity 59.9%; Pred. No. 3.1e-83;
Matches 214; Conservative 53; Mismatches 79; Indels 11; Gaps 5;

QY 17 EFMVQNHADVSCVGMFFVLGLMFEFTAEMSIVFLTLQHGCVVPA-EGLPSSGRTLYHYG 75
DB 1 EFLVQNHADIVSCVAMVFLGLMFEITAKVSIIFVLQYVNTLPATYEQNTESAFLYYYG 60
QY 76 VKDLATVFFVYMLVAIIHATIQEYVLDKLSRRQLTKGKQNKLEAGOLSVFVIVSGIWG 135
DB 61 IKDLATVFFVYMLVAIIHATIQEYVLDKLNRRMHFSKTKHKNESQLSAFYLFSCIWG 120
QY 136 MIILASENCSDPTLLWKSPHNMFTQMKFFYISQLATWFHSPFELYQKVRQDIPGQ 195
DB 121 TFLISENISTDPTLLWRAPPHNLMTFQMKFFYISQLATWFHAFPELYFQTKTKEDIPRQ 180
QY 196 LIYIGLHLPHIGAYLLYLNHLGLLLMLHYAVELLSVCSLLYFGDERYOKGLSLWPV 255
DB 181 LVYIGLYLPHIAGAYLLNHLGLVLLVLYHVFELFHSRLYFFDEKQKQGLSLWAVL 240
QY 256 FTSGRLVTLIVSVVTVGLHLA-GTNRGNALSGNVNLAALAVLSSSCSIQYIITWTET 314
DB 241 FVLGRLLTLILSVLVGVGLARAENKLPSTGNENVLAVRIAVLASICITQAFMMWKEI 300
QY 315 TWVLQRLWLEDANLHVCGRRK-----RSRS-RKGTENGVE----NPNRIDSPPKKKA 362
DB 301 NFQRLRWREHSAPQAPAVKKKPPVTGKRSXXKGTENGVTGTSNGADSPNRNKRKS 357
```

```

RESULT 4
Y557_HUMAN STANDARD; PRT; 370 AA.
ID Q15035;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein KIAA0057.
GN KIAA0057.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow; PubMed=7584044;
RX MEDLINE=96051198; Miyajima N., Suzuki T., Tanaka A., Sato S.,
RA Nomura N., Nagase T., Ishikawa K.-I., Tabata S.,
RA Seki N., Kawarabayashi Y.,
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 1:223-229(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Tracey A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE LASS1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D31762; BAA06540.1; -
DR EMBL; AL049611; CAB71119.1; -
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 23 43 POTENTIAL.
FT TRANSMEM 76 96 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 251 271 POTENTIAL.
FT TRANSMEM 288 308 POTENTIAL.
SQ SEQUENCE 370 AA; 43327 MW; 9B5183F1A3D45366 CRC64;

Query Match 39.9%; Score 754; DB 1; Length 370;
Best Local Similarity 42.4%; Pred. No. 1.1e-55;
Matches 158; Conservative 63; Mismatches 132; Indels 20; Gaps 5;

QY 5 KKNARPPVLSSHFMVQNHADVSCVGMFFVLGLMFEFTAEMSIVFLTLQHGCVVPAEGL 64
DB 4 RRTKSTPLSFSEFVHNHADIGFCLVLCVLGLMFEFTAEMTAKTFLFLPQINISVPT- 60
QY 65 PSCSRTL-YHYGVKDLATVFFVYMLVAIIHATIQEYVLDKLSRRQLTKGKQNKLEAGQ 123
DB 61 -ADSEVHYHYGPKDLVTILFYITILHVVQVEYILDKISRLHLSKVHKSFNESGQ 119
QY 124 LSVFYIVSGTWGMIILASENCSDPTLLWKSPHNMFTQMKFFYISQLATWFHSPFELY 183
DB 120 LVVFHETSIVCWCFVYVVTGTYLTPRSLWEDYFVHLPFQVKKFYLCQLAYLHALPELY 179
QY 184 FQVVRQDIPGQIYICLHLPHIGAYLLYLNHLGLLLMLHYAVELLSVCSLLYFGDE 243
DB 180 FQVVRKEEIPRQYICLYLVHIAGAYLLNLSRLGLLILLQLQYSTEFHTARLFYFADE 239
QY 244 RYQKGLSLWPVIVTSGRLVTLIVSVVTVGLHLA-GTNRGNALSGNVNLAALAVLSSS 302
```

FT	173	174	ML -> IV (IN REF. 2).
FT	220	220	F -> C (IN REF. 2).
FT	301	411	VTFVFGFLVFFWIVLHVVRINILMSVLTEPFRHEGNCVL
FT			NFAAQYQKWCTLSLPIFVLAALQVNLVYLILRLYL
FT			IWOQIKDRSDSDSAENESKECE -> TEISGIWE
FT			KQEDSNNDPTALRSPNETSKOVKPDLLVLPNTNRNAL
FT			LEAKSRVPTAITDTEPSVLVYPIPGDNDLSRSLVNFLL
FT			GVLAGARQGLQNRLARNEK (IN REF. 1).
SQ	411 AA:	48454 MW:	91676D56AC053FC CRC64:

[illegible]

```

RESULT      6
YKAB_YEAST
ID          YKAB_YEAST      STANDARD;      PRT;      418 AA.
AC          P28496;
DT          01-DEC-1992 (Rel. 24, Created)
DT          01-JUN-1994 (Rel. 29, Last sequence update)
DT          16-OCT-2001 (Rel. 40, Last annotation update)
DE          Hypothetical 49.0 kDa protein in UFD4-CAP1 intergenic region.
GN          YKL008C OR YKL156.
OS          Saccharomycetaceae (Baker's yeast).
OC          Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC          Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN          [1]
RN          RN
RP          SEQUENCE FROM N.A.
RX          MEDLINE=93255906; PubMed=8488728;
RA          Boyer J., Pascolo S., Richard G.F., Dujon B.;
RT          "Sequence of a 7.8 kb segment on the left arm of yeast chromosome XI
RT          reveals four open reading frames, including the CAP1 gene, an intron-
RT          containing gene and a gene encoding a homolog to the mammalian UOG-1
RT          gene."
RT          Yeast 9:279-287(1993).
RL          [2]
RP          SEQUENCE OF 1-149 FROM N.A.
RP          MEDLINE=93077675; PubMed=1447293;
RR          Amatruda J.F., Gattermeir D.J., Karpova T.S., Cooper J.A.;
RX          "Effects of null mutations and overexpression of capping protein on
RT          morphogenesis, actin distribution and polarized secretion in yeast.";
RT          J. Cell Biol. 119:1151-1162(1992).
RL          J. Cell Biol. 119:1151-1162(1992).
CC          -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC          -!- SIMILARITY: BELONGS TO THE LASS1 FAMILY.
CC          This SWISS-PROT entry is covright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X61398; CAA43670.1; -;
DR EMBL; S59773; AAC60549.1; -;
DR EMBL; Z28008; CAA81843.1; -;
DR PIR; S30134; S30134;
DR SGD; S0001491; YKL008C;
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 173 193 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 356 376 POTENTIAL.
SQ SEQUENCE 418 AA; 48992 MW; 7691BAG23AC0460A CRC64;

Query Match 7.2%; Score 137; DB 1; Length 418;
Best Local Similarity 23.5%; Pred. No. 0.00034;
Matches 61; Conservative 44; Mismatches 93; Indels 62; Gaps 11;

QY 67 GSRTLYHYGVKDLATVFFYVMAIIHATIQEYVLDKLSRRQLTKGQKNEAGLSV 126
DB 121 GDTNAYGKGLNCLCFVYYIFFFLREFLMDVVRPFAIRLHVTSKHKRIKMEQYAI 180
QY 127 FYI-VSGINGMIL-ASENCLSDPTLLWKSQPHNMFTQMKFFYISOLAYV----- 175
DB 181 FYTGVSPGFIYCMYHSDLWFNTKAMRYTPDFTNPFKVFYLGQAFAWAQACILVL 240
QY 176 -----FHSPPEL-----YFQVKRKQDIPGQLIYIGLHLFHGGAYLLYNH 216
DB 241 QLEKPRKHNEHFTFHVITLLLSWTSYVFTFKMGLP---IYITMDVDFLSFKTLNY 297
QY 217 L--GLL-----LLMLHYA-VELLSSVCS-----LLYFGDERYQKGLSLWPI 254
DB 298 LDSGLAFFFAIFVAVIYLRHVIKILMSVLTPQRTGNYVFNATQYKWCWISL-PI 356
QY 255 VFISGRVLTVSVTVGLH 274
DB 357 VFV-----LIGALQLVNLV 370

RESULT 7
ID YHXF_SCHPO STANDARD; PRT; 384 AA.
AC O59735;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 42.4 kDa protein C3E17.15c in chromosome II.
GN SPB3E7.15C OR SPB4F6.02C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lyne M., Wood V., Rajandream M.A., Barrell B.G., Brown D.,
RA Churcher C.M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Gwilliam R., Rajandream M.A., Barrell B.G., Skelton J., Churcher C.M.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE LASSI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AL023534; CAA19018.2; -;
DR EMBL; AL031534; CAA20722.2; -;
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 195 215 POTENTIAL.
FT TRANSMEM 233 253 POTENTIAL.
FT TRANSMEM 270 290 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
SQ SEQUENCE 384 AA; 45335 MW; 82FCF8EA6638849A CRC64;

Query Match 7.0%; Score 133; DB 1; Length 384;
Best Local Similarity 20.5%; Pred. No. 0.00067;
Matches 48; Conservative 45; Mismatches 117; Indels 24; Gaps 7;

QY 41 EGTAEMSIIVLT-----OHGVVVPAGEPLSGSRTLYHYGVKDLATVFFVML 87
DB 56 EKTWVPLILLTLVGWYFVNPNGYIKYIGFL-SYIPGTNPAGYKGRDLIAFLFYAL 114
QY 88 VAIITHATIQEYVLDKLSRRQL--TKGQKNEAGLSVFIYVSGIWGMIL-ASENCL 145
DB 115 FFFCFREFIMQEIIRIGRHNIRAPAKLRFEQATCLYFTVMSGLYVMKQTPWF 174
QY 146 SDPTLLWKSQPHNMFTQMKFFYISOLAYWHPSPPELYFQ--KVRKQDIPGQLIYIGLHL 203
DB 175 FNTDAFWEEYFHFYHVGVSFAFYLIAEAYWTOQALVILQLEKPRK-DFRELVVHHITL 233
QY 204 FHGGAYLLNLHGLLLMLHYAVELLSSVCSLLYFGDERYQKGLSLWPIVFI 257
DB 234 LLIGLSYFHTWIGLAVFITMDTSDIWLALSKCL-----NYVNTVIVIPFVI 282

RESULT 8
ID LAG1_SCHPO STANDARD; PRT; 390 AA.
AC P78970; O13860;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Longevity-assurance protein 1 (Longevity assurance factor 1).
GN LAG1 OR SPAC1A6.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Chanda E.R., Lingner C., Ko Z., Young P.G.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V., Wood V.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE AGING PROCESS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE LASSI FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

[illegible]

Db 102 SGDWTLPDLA-----WLNTP-----VIGAKLYAGV 127
QY 202 H-LFHIGGAYLL-----YLN-----HLGLLLMLHYAVELSSVCSLLYFGDER 244
Db 128 HNLDMGCTAIMAKVRPIYTTTTTWFVGOAHIG--RFVHCALMLLFS--ALLYMRGEQ 183
QY 245 YQKG-----LSLWPIFIVS-GRVTLIVSVVTVGLHLAGLNRNGNAL 285
Db 184 VAQCIIRHFATRLAGVRGDAVLLAAQAIRAVLGVVVVTVLQVAVLGGI-----GLAV 235
QY 286 SGNNVNLAAKIVLSSCSIQ-----VYIWTWLTWV 317
Db 236 SG-VPRATLLTVMILSCLVOLGPLPVLPAITWLYWTGDTTW 277

RESULT 13
SOAL_MOUSE
ID SOAL_MOUSE STANDARD; PRT; 540 AA.
AC Q61263; Q64180;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Cholesterol acyltransferase
1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).
GN SOAT1 OR ACAT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=96064687; PubMed=7592824;
RA Uelman P.J., Oka K., Sullivan M.C., Chang T.-Y., Chang C.C.Y.,
Chan L.;
RT Tissue-specific expression and cholesterol regulation of acylcoenzyme
A:cholesterol acyltransferase (ACAT) in mice. Molecular cloning of
mouse ACAT cDNA, chromosomal localization, and regulation of ACAT in
vivo and in vitro.;
RL J. Biol. Chem. 270:26192-26201(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96158986; PubMed=8579615;
RA Green S., Steinberg D., Quehenberger O.;
RT "Cloning and expression in Xenopus oocytes of a mouse homologue of the
human acylcoenzyme A: cholesterol acyltransferase and its potential
role in metabolism of oxidized LDL";
RL Biochem. Biophys. Res. Commun. 218:924-929(1996).
CC -!- FUNCTION: CATALYZES THE FORMATION OF FATTY ACID-CHOLESTEROL
ESTERS. PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY
CHOLESTEROL ABSORPTION.
CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol
ester.
CC -!- SUBCELLULAR LOCATION: Integral membrané protein. Endoplasmic
reticulum.
CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; L42293; AAC42075.1; -;
DR EMBL; S81092; AAB36050.1; -;
DR MGD; MGI:104665; Soat1.
DR InterPro; IPR002688; ACAT.
DR Pfam; PF01800; ACAT; 1.
KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;
KW Cholesterol metabolism.
FT TRANSMEM 132 152
POTENTIAL.

FT TRANSMEM 311 331
FT TRANSMEM 352 372
FT TRANSMEM 460 480
FT TRANSMEM 488 508
FT TRANSMEM 495 515
SQ SEQUENCE 540 AA; 63739 MW; 8EF900C8BCDF73C0 CRC64;
Query Match 5.0%; Score 94.5; DB 1; Length 540;
Best Local Similarity 21.5%; Pred No 1.6;
Matches 45; Conservative 34; Mismatches 93; Indels 37; Gaps 6;
QY 48 IVFELTQHGVS-----VPAEGLPSGSRSLTYHYGVKDLATVFFYMLVAIIHATIQEYV 100
Db 215 LLFLVFGVGLGVFVPTVVLAYLTPPASR-----FILLEQIRLIMKAHSFV 261
QY 101 LDKLSRLQLTKGQNK--LNEAGOLSVFYIVSGINGMILASENCISDPTLLKWSQPHN 158
Db 262 RENIPRVLNAAKEKSSKDPFTVNOYLYF-----LFAPTLYRDNYPRTPVW---GYV 313
QY 159 MMTFOMKFFVYSOLAYWFHSFPPELYFOKVRKODIPGOLYIGLHFGHGGAYLLYLNHLG 218
Db 314 AMQFLQVFGCLFYVYVIFERLCAPLFRNIQKQEPFSARVILCVFNSILPGVLIILFSLFFA 373
QY 219 LLMLHLYAVVELSSVCSLLYFGDERYOK 247
Db 374 FLHCLWLNFAEMLR-----FGDRMFYK 395
RESULT 14
ID COXM_SULAC STANDARD; PRT; 788 AA.
AC P39481;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Quinol oxidase polypeptide I/III (EC 1.9.3.-).
GN SOXM.
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RX MEDLINE=94357214; PubMed=8076636;
RA Luebben M., Arnaud S., Castresana J., Warne A., Albracht S.P.J.,
Saraste M.;
RT "A second terminal oxidase in Sulfolobus acidocaldarius";
RL Eur. J. Biochem. 224:151-159(1994).
CC -!- FUNCTION: TERMINAL OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SOXM
FORMS THE FUNCTIONAL CORE OF THE ENZYME COMPLEX.
CC -!- FUNCTION: CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING
IN A QUINOL ARE TRANSFERRED TO THE BIMETALLIC CENTER OF SOXM
FORMED BY A HEME AND COPPER B.
CC -!- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
c + 2 H(2)O.
CC -!- COFACTOR: TWO HEME GROUPS AND COPPER B.
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -!- SUBUNIT: FORMS A COMPLEX WITH AT LEAST SOXM AND A 30 kDa RIESKE
FE-S PROTEIN, BUT NEITHER WITH SOXA NOR SOXB.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- PTM: SOXM IS PROBABLY A PRECURSOR FORM OF SUBUNIT I AND III.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE HEME-COPPER
RESPIRATORY OXIDASE FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE CYTOCHROME C
OXIDASE SUBUNIT 3 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC DR EMBL; X73567; CAA51969.1; -.
CC DR HSSP; P00396; 10CC.
CC DR InterPro; IPR000883; COX1.
CC DR Pfam; PF00115; COX1; 1.
CC DR Pfam; PF00510; COX3; 1.
CC DR PRINTS; PR01165; CYCOXIDASE1.
CC DR PROSITE; PS00077; COX1; 1.
CC DR PROSITE; PS0253; COX3; 1.
CC KW Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain.
CC FT TRANSMEM 19 39 POTENTIAL.
CC FT TRANSMEM 58 78 POTENTIAL.
CC FT TRANSMEM 100 119 POTENTIAL.
CC FT TRANSMEM 152 169 POTENTIAL.
CC FT TRANSMEM 187 206 POTENTIAL.
CC FT TRANSMEM 232 252 POTENTIAL.
CC FT TRANSMEM 276 294 POTENTIAL.
CC FT TRANSMEM 307 326 POTENTIAL.
CC FT TRANSMEM 337 356 POTENTIAL.
CC FT TRANSMEM 369 388 POTENTIAL.
CC FT TRANSMEM 411 432 POTENTIAL.
CC FT TRANSMEM 449 468 POTENTIAL.
CC FT TRANSMEM 526 544 POTENTIAL.
CC FT TRANSMEM 547 564 POTENTIAL.
CC FT TRANSMEM 614 634 POTENTIAL.
CC FT TRANSMEM 657 676 POTENTIAL.
CC FT TRANSMEM 690 707 POTENTIAL.
CC FT TRANSMEM 729 749 POTENTIAL.
CC FT TRANSMEM 766 786 POTENTIAL.
CC FT METAL 62 62 IRON (HEME A) (PROBABLE).
CC FT METAL 238 238 COPPER B (PROBABLE).
CC FT METAL 242 242 COPPER B (PROBABLE).
CC FT METAL 288 288 COPPER B (PROBABLE).
CC FT METAL 289 289 COPPER B (PROBABLE).
CC FT METAL 374 374 IRON (HEME A(3)) (PROBABLE).
CC FT METAL 376 376 IRON (HEME A) (PROBABLE).
CC SQ SEQUENCE 788 AA; 87082 MW; 3939C16CDB8A08AD CRC64;

Query Match 5.0%; Score 94.5; DB 1; Length 788;
Best Local Similarity 19.5%; Pred. No. 2.5;
Matches 73; Conservative 57; Mismatches 98; Indels 147; Gaps 19;

QY 28 SCVG-MFVVLGL--MFGETAEMSIV-----FLTQ-----HG-----VVVP-AE 62
DB 17 SDVGQMYIVLGIVALLIGSVNAALIRDOLFNNLNAVDDYDAVTLHGIFMFFVYVMP 76
QY 63 GLPS-----GSRTLYHYGVKDLATVFFYMLVAIIHATIQE-----YVLDKLSR 106
DB 77 GFANYLVPRMIGADHLYWPKINALS---FWMLVPAVILAALSPLLGAVDLGWYMAPLSV 133
QY 107 RLQLTGKGNKNEAGQLSVFYIVSG-----IWGMII 138
DB 134 ETTVNYGLGTNL-----IQIALISGLSSTLTGVNFVMTIKMKVPYLYKMPFLVWGFF 188
QY 139 LASENCLSDPILLWKSQPHNMFTQMKFFYISQL-----AYWHSF 179
DB 189 TAILMIAMPSTL-----TAGLVFAYLERLWGTFFDSALGGSPVLVWQQLFWFFGH 238
QY 180 PELY-----FQKVRKODIPQQLIYIGLHFLHFGAYL---LYNLHLGLLL 222
DB 239 PEVYIILIPAMGLYSELFPKARREIFG---YTAIALSSIAIALSALGVWVHH--MFTA 293
QY 223 MLHYAVELLSSVCSL-----LYFGDERYQKGLS---WPIVFTSGRLVT 263
DB 294 IDNTLVQIVSSATTMAIAPSGVKVNLNWTATLYGGEIRYKPTILLISFIVMFLGGITG 353
QY 264 LIVSVVTVGLHLAGT 278
DB 354 VFFPLVPIDYALNGT 368
```

```
RESULT 15
ATP6_BUCAP
ID ATP6_BUCAP STANDARD; PRT; 272 AA.
AC O51878;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ATP synthase A chain (EC 3.6.3.14) (Protein 6).
GN ATPB.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=97361981; PubMed=9216881;
Clark M.A., Baumann P.;
"The (F1F0) ATP synthase of Buchnera aphidicola (endosymbiont of aphids): genetic analysis of the putative ATP operon.";
Curr. Microbiol. 35:84-89(1997).
[2]
SEQUENCE FROM N.A.
MEDLINE=98184963; PubMed=9516544;
Clark M.A., Baumann L., Baumann P.;
"Sequence analysis of a 34.7-kb DNA segment from the genome of Buchnera aphidicola (endosymbiont of aphids) containing groEL, dnaA, the atp operon, gidA, and rho.";
Curr. Microbiol. 36:158-163(1998).
CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE (BY SIMILARITY).
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0) HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
```

```
EMBL; AF008210; AAC38116.1; -.
HSSP; P00855; 1C17.
InterPro; IPR000568; ATP_synt_A.
Pfam; PF00119; ATP_synt_A; 1.
PROSITE; PS00449; ATPASE_A; 1.
KW Hydrogen ion transport; Cf(0); Transmembrane.
FT TRANSMEM 41 61 POTENTIAL.
FT TRANSMEM 110 130 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
FT TRANSMEM 222 242 POTENTIAL.
FT TRANSMEM 243 263 POTENTIAL.
SQ SEQUENCE 272 AA; 31253 MW; DCD8C7D2C98C37CC CRC64;

Query Match 5.0%; Score 94; DB 1; Length 272;
Best Local Similarity 23.5%; Pred. No. 0.8;
Matches 65; Conservative 45; Mismatches 87; Indels 80; Gaps 16;

QY 73 HYGKDLATVFFYMLVAIIHATIQEYVLDKLSRLQLTKGKQNKNEAGQLSVFYIVSG 132
DB 38 HFVNLNDSIIFSLVLCFFLSIF--YTVAK-----KITTVGNPGLQASIELIDFIRSN 90
QY 133 IWGMIIASENCLSDP-----TLLWKSQPHNMFTQMKFF-YISQLAYWFHSFPELYQKV 187
DB 91 VKSM--YQGKNPLIAPLSLTVFVWVFLNMLMDLIPIDFFPFISE--RFFH-FPAM----- 140
```

```
Qy 188 RKQDIPGQLIYIGLHFIHGAYLLYLNHLGLLLMLHYAVEL--LSSVCSLLYFGDERY 245
Db 141 --RIVPSADINITYLSM-----SLGVFILLIFYSVKMKGLIGFC----- 176
Qy 246 QKGLSL-----WPIVFISG---RLVTLIVSVTVVGLHLAGTNRGNALSGN----- 288
Db 177 -KELTQPFNHPVFFIFNFLLELVSLSKPISLGLRLF-----GNMYSGEMIFILIAGLL 230
Qy 289 -----VNVLAAKIAVLSSSCSTIOVYITWTLTVWL 318
Db 231 PWMSQFFLNVPWAIFHL--IISLQAFIFMVLTVYL 265
```

Search completed: September 6, 2002, 17:21:14
Job time: 632 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 17:09:51 ; Search time 80.19 seconds
(without alignments)
783.105 Million cell updates/sec

Title: US-09-807-470-2
Perfect score: 1892
Sequence: 1 MGLKKNARNPPVLSHEFMV.....NGVENPNRIDSPPKKKEKAP 363

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_19:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Description
1	1125.5	59.5	Q91V04 mus musculus
2	1102	58.2	Q90ZM0 xenopus lae
3	1024	54.1	Q90ZM1 brachydanio
4	762	40.3	Q90ZL9 xenopus lae
5	727	38.4	Q924Z5 mus musculus
6	566	29.9	Q9W5C2 drosophila
7	566	28.9	Q9W5C3 drosophila
8	561	29.7	Q9U1L3 drosophila
9	441.5	23.3	Q9U3P5 caenorhabdi
10	434.5	23.0	Q9XXK7 caenorhabdi
11	348	18.4	Q9CVJ6 mus musculus
12	163.5	8.6	Q9D6J1 mus musculus
13	138.5	7.3	Q9HA82 homo sapien
14	131	6.9	Q9M6A4 lycopersico
15	119.5	6.3	Q95RN6 drosophila
16	116.5	6.2	Q9CHC2 lactococcus

17	115.5	6.1	296	10	Q9LJK3	Q91jk3 arabidopsis
18	115.5	5.8	297	10	Q9M6A2	Q9m6a2 arabidopsis
19	110	5.8	380	4	Q96G23	Q96g23 homo sapien
20	109	5.8	380	11	Q924Z4	Q924z4 mus musculus
21	107.5	5.7	614	16	Q31615	Q31615 bacillus su
22	106.5	5.6	639	16	Q98KL8	Q98kl8 rhizobium l
23	105.5	5.6	310	10	Q9LDF2	Q9ldf2 arabidopsis
24	103.5	5.5	393	16	Q97LW0	Q97lw0 clostridium
25	103	5.4	414	11	Q9D6K9	Q9d6k9 mus musculus
26	103	5.4	414	11	Q924Z3	Q924z3 mus musculus
27	102.5	5.4	522	16	Q9K9F0	Q9k9f0 bacillus ha
28	101.5	5.4	397	2	O05362	O05362 actinobacil
29	101	5.3	308	10	Q9FZ69	Q9fz69 arabidopsis
30	100.5	5.3	360	5	Q9XWE9	Q9xwe9 caenorhabdi
31	100.5	5.3	397	16	O84825	O84825 chlamydia t
32	100	5.3	354	16	Q9CLS8	Q9cls8 pasteurella
33	100	5.3	406	13	Q90YI6	Q90yi6 brachydanio
34	99.5	5.3	395	16	Q9PLA2	Q9pla2 chlamydia m
35	99	5.2	286	2	Q9X6B4	Q9x6b4 yersinia pe
36	98.5	5.2	441	13	Q91I91	Q91i91 brachydanio
37	98.5	5.2	460	2	Q9XD16	Q9xd16 leptospira
38	98.5	5.2	722	16	Q9KG28	Q9kg28 bacillus ha
39	98	5.2	232	5	Q9VXF0	Q9vxf0 drosophila
40	97.5	5.2	397	2	O6G245	O6g245 actinobacil
41	96	5.1	378	5	Q9Y1V7	Q9y1v7 halocynthia
42	95.5	5.0	324	2	Q9ZIM4	Q9zim4 listeria mo
43	95.5	5.0	411	17	Q980X8	Q980x8 sulfolobus
44	95.5	5.0	441	13	Q90ZE4	Q90ze4 brachydanio
45	95.5	5.0	751	10	Q93XX4	Q93xx4 arabidopsis

ALIGNMENTS

RESULT 1

ID	Q91V04	PRELIMINARY;	PRT;	374 AA.
AC	Q91V04;			
DC	01-DEC-2001 (Tremblrel. 19, Created)			
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)			
DE	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE	TRAM1 (UNKNOWN) (PROTEIN FOR MGC:11724).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Hartmann E.;			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Strausberg R.;			
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AY029764; AAK38167.1; -			
DR	EMBL; BC012401; AAH12401.1; -			
SQ	SEQUENCE 374 AA; 43039 MW; E6C65250F68E4393 CRC64;			
Query Match 59.5%; Score 1125.5; DB 11; Length 374;				
Best Local Similarity 58.4%; Pred. No. 5.6e-93;				
Matches 218; Conservative 58; Mismatches 86; Indels 11; Gaps 4;				
QY	1	MGLKKNARNPPVLSHEFMVQNHADMVSCVGMFFVLGLMFEGETAEMKSIVELTLQHGWWVP	60	
DB	1	MAIRKSNKNPPVLSHEFLQNHADIVSCLMLFLGLMFEVAKGAIIFVALQYNNVTRP	60	
QY	61	A-EGLPGSGRTLYHYGVKDLATVFFYMLVAIIITHATIQEYVLDKLSRRLQTLTKGKNKLN	119	
DB	61	ATEQATEASLYHYGKDLATVFLYMLVAIIITHATIQEYVLDKLSRRLQTLTKGKNKLN	120	
QY	120	EAGQLSVFIVSGIWGMIIILASENCLSDPTLLWKSPHNMTQMKFFYISQLAYFHFSF	179	
DB	120	EAQLSVFIVSGIWGMIIILASENCLSDPTLLWKSPHNMTQMKFFYISQLAYFHFSF	179	

```
Db 121 ESGLSAFYLFACVNGTIFILISENYISDPTILWRAYPHNLMTFQTKFFYISQLAYWLHAF 180
QY 180 PELYFQVRKODIPGOLIIYIGLHFIHGAYLLYNHLGLLMLHYAVELLSSVCSLLY 239
Db 181 PELYFQTKKEDIPQLVYIGLYLFIHGAYLLNHLGLVLLVLYHVFELFHSRLFY 240
QY 240 FDERYQKGLSWPIVFTSGRLVTVIVSVTVVGLHLA-GTNRNGNALSGNVNVLAAKIAV 298
Db 241 FSDEYQKGFSLWAVFLVGLRLLTLLSVLVGFLGARAENQKLDFTGTFNVLAVRIAV 300
QY 299 LSSSCSIQVIITWLTWMLQWLEDANLHVCGRKR-----SRKKGTEGVG---NP 349
Db 301 LASICITQAFMMWKFINQLRRWREHSAFAQPPVKRPVAVTKGRSSRKGTENGNGVTISN 360
QY 350 NRIDSPPKKKEKA 362
Db 361 NGADSPRRKKEK 373

RESULT 2
ID Q902M0 PRELIMINARY; PRT; 373 AA.
AC Q902M0;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE TRAM1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Hartmann E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY029534; AAK40302.1; -.
SQ SEQUENCE 373 AA; 43121 MW; 62AFA9E9859769DF CRC64;
```

```
Query Match 58.2%; Score 1102; DB 13; Length 373;
Best Local Similarity 56.1%; Pred. No. 7.2e-91;
Matches 208; Conservative 64; Mismatches 89; Indels 10; Gaps 3;

QY 1 MGLRKNARNPPVLISHEFWONHADMVSCVGMFFVLGLMFEGETAEMSVFTLQHGYYVP 60
Db 1 MGLRKNKSTPPVLSHEFIQNHADIVSCVAMVFLGLMFEITSKVAVLFTVQYNTVP 60
QY 61 AGLPSGSTRLLY-HYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRLQLTGKONKLN 120
Db 61 VEGVLGEPTSLYHYGDKDMATVFFYMLVAIIHATIQEYVLDKLNRRMHFSKTHSKFNE 120
QY 121 AGQLSVFYIVSGIWMILASENCLSDPTILWKSQPHNMFTQMKFFYISOLAYWFHSP 180
Db 121 SQGLSAFYLFSCIGWASIVSENTFSDPSLWKGYPHTFPFQMKFFYISOLAYWFHAF 180
QY 181 ELYFQVRKODIPGOLIIYIGLHFIHGAYLLYNHLGLLMLHYAVELLSSVCSLLYF 240
Db 181 ELYFQTKKEDIPQLVYIGLYLFIHGAYLLNHLGLVLLVLYHVFELFHSRLFY 240
QY 241 GDERYQKGLSWPIVFTSGRLVTVIVSVTVVGLHLA-GTNRNGNALSGNVNVLAAKIAVL 299
Db 241 SNERYQKGFVWAVFLVGLRLLTLLSVLVGFLGARAENQKLDLSNGNFILAIRITVL 300
QY 300 SSSCSIQVIITWLTWMLQWLEDANLHVCGRKR-----TENGVE---NPN 350
Db 301 ASICITQAFMMWKFINQLRRWREHSSQPSSQRKATSAKGRKENGNGVTISN 360
QY 351 RIDSPPKKKEK 361
Db 361 GADSPRRKKEK 371
```

```
RESULT 3
Q90ZM1
ID Q90ZM1 PRELIMINARY; PRT; 369 AA.
AC Q90ZM1;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE TRAM1.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Hartmann E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY029529; AAK40297.1; -.
SQ SEQUENCE 369 AA; 41743 MW; A3EE6263E3165846 CRC64;
```

```
Query Match 54.1%; Score 1024; DB 13; Length 369;
Best Local Similarity 52.4%; Pred. No. 7.2e-84;
Matches 193; Conservative 67; Mismatches 102; Indels 6; Gaps 4;

QY 1 MGLRKNARNPPVLISHEFWONHADMVSCVGMFFVLGLMFEGETAEMSVFTLQHGYYVP 60
Db 1 MGLRKNKSTPPVLSHEFIQNHADIVSCVAMVFLGLMFEITSKVAVLFTVQYNTVP 60
QY 61 AGLPSGSTRLLY-HYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRLQLTGKONKLN 119
Db -61 ANGGEAEATVNYFHYGLKDVATIFFYMLVAIIHATIQEYVLDKLNKMHFSKTHSKFN 120
QY 120 BAGQLSVFYIVSGIWMILASENCLSDPTILWKSQPHNMFTQMKFFYISOLAYWFHSP 179
Db 121 ESGLSAFYLFSCIGWASIVSENTFSDPSLWKGYPHTFPFQMKFFYICOLGYWLHAI 180
QY 180 PELYFQVRKODIPGOLIIYIGLHFIHGAYLLYNHLGLLMLHYAVELLSSVCSLLY 239
Db 181 PELYFQKAKKEDIPQLVYISLVVHAGAYALNLRGLVLLVLYHVFELIPHSRLIY 240
QY 240 FDERYQKGLSWPIVFTSGRLVTVIVSVTVVGLHLAAGTNRNG-NALSGNVNVLAAKIAV 298
Db 241 FSNBERQSGTVAWFLVGLRLLTLLSVLVGFLGAEKQGLNLAESFNVLFRVTV 300
QY 299 LSSSCSIQVIITWLTWMLQWLEDANLHVCGRK---RRSRKGT-ENGVENPNRIDS 354
Db 301 LAAICITQAFMMWKFINQLRRWREQAOTLTKKSSSKSKRANGVSGVSGAGADS 360
QY 355 PPKKKEKA 362
Db 361 PRARKEKS 368
```

```
RESULT 4
Q90ZL9
ID Q90ZL9 PRELIMINARY; PRT; 371 AA.
AC Q90ZL9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE TRAM2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Hartmann E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY029535; AAK40303.1; -.
SQ SEQUENCE 371 AA; 43548 MW; 582231A5286D130B CRC64;
```

Query Match	40.3%;	Score 762;	DB 13;	Length 371;
Best Local Similarity	41.2%;	Pred. No. 2.5e-60;		
Matches 159;	Conservative 67;	Mismatches 122;	Indels 38;	Gaps 7;
QY	1	MGLRKKNARNPPVLSHEFMVQNHADMVSCVGMFFVL-----GLMFEGTAEMSIIVFLTLQHG	56	
Db	1	MAFERR--KSYPLFSQEFVIHNDAD-----IGFFVLCVLIGLMFEVTAFTAFILPQYN	54	
QY	57	VVVPABGLPSGRTLYHYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRQLTKGKN	116	
Db	55	SSIQT---LDGEILYHYGVKDLVTLFVYVIAIIHLAIHQEYILDKINKRLHLSKVQS	111	
QY	117	KNEAGOLSVFYIVSGIWMIIASENCLSDDPTLLWKSQPHNMFTOMKFFYISOLAYWF	176	
Db	112	RNENSGQLAFAHLSMFWCLYSATSEGLYSYPKTWESYPHYLPQVQVFFVLCQAYWL	171	
QY	177	HSFPPELYFQVRKQDIPGQIIYIGLHFIHGAYLLYLNLHLGLLLMLHYAVELLSSVCS	236	
Db	172	HALPELYFQVRKKEEVPRLQYIVLYLLHAGAYLLNLTLGLILLILQSVAEFLFIAR	231	
QY	237	LIYFGDERYQKGLSWPIVIFISGRIVTLIVSVTVGLHLAGTN--RNGNALSGNVNVLAK	295	
Db	232	LFYFTDENNRQLFNAGVVFVITRLFTLTLSVITLFGFLARAEVHTFPDKGTNLTLFR	291	
QY	296	IAVLSSCSIQVYITWTLTTVWLQWLELDANLHVCKRRRSRKGK-----	342	
Db	292	MVLLLMCVSQTWMMWRFHFLQLRWRE-----CCKEQAAKRKSVAVAMKQAKVIKR	345	
QY	343	-----ENGVENPNRIDSPPKKKEKAP	363	
Db	346	ESGYHENGVVKAENGSTPQKKIKSP	371	
RESULT	5			
Q92425				
ID	Q92425	PRELIMINARY;	PRT;	370 AA.
AC	Q92425;			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	TRAM2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Hartmann E.;			
RL	Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL: AY029530; AAK40298.1; --			
SQ	SEQUENCE 370 AA; 43182 MW; F8E768AFB0582548 CRC64;			
Query Match	38.4%;	Score 727;	DB 11;	Length 370;
Best Local Similarity	40.7%;	Pred. No. 3.4e-57;		
Matches 151;	Conservative 70;	Mismatches 134;	Indels 16;	Gaps 5;
QY	5	KNANRNPVLSHEFMVQNHADMVSCVGMFFVYGLMFEGTAEMSIIVFLTLQHG	VVVPABGL	64
Db	4	RRRTKSYPLFSQEFIIHNDADIGFCLVLCVLIGLMFEVTAFTAFILPQYNISVPT---	60	
QY	65	PSGSRTL-YHYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRQLTKGKNLEAGQ	123	
Db	61	-ADSETVHYHGVKDLVTLFVWVITFIHFAVYIYLDKLSRHLKSRHKSFTNESQ	119	
QY	124	LSVFYIVSGIWMIIASENCLSDDPTLLWKSQPHNMFTOMKFFYISOLAYWFHSPPELY	183	
Db	120	LIVFHLSAVAMCFYIVTVEGTYLTPRSLWEDYPHYLFSQVAFVYLGQAYLWLSLPELY	179	
QY	184	FQVRKQDIPGQIIYIGLHFIHGAYLLYLNLHLGLLLMLHYAVELLSSVCSLLYFGDE	243	

Db	180	FQVRKEEVPRLQYICLYLLHITGAYLLNLSRGLILLLOYSYFALFHWARLHFPADE	239
QY	244	RYQKGLSWPIVIFISGRIVTLIVSVVTVGLHLAGT--NRNGNALSGNVNVAAKIYVLSSS	302
Db	240	NNERLFNAAVAVGFTVRLFTLTAVLTIGFGLARVENQVDFPEKGNFTLPCRLGMLLV	299
QY	303	CSIQVYITWTLTTVMQLQWLE-----DANLHVCG-----RKRSRSRKGTGVENPNRI	352
Db	300	CVAQAWLMWRFTHSQRHWRKWEQSAKRRVSAPRPPAKLLKREPFGYHENG	359
QY	353	DSPPKKKKKAP	363
Db	360	TSSRTKKLKSP	370
RESULT	6		
Q9W5C2			
ID	Q9W5C2	PRELIMINARY;	PRT: 368 AA.
AC	Q9W5C2;		
DT	01-MAY-2000 (TREMBlrel. 13, Created)		
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)		
DT	01-MAR-2001 (TREMBlrel. 16, Last annotation update)		
DE	EG:BACR7A4.5 PROTEIN.		
GN	EG:BACR7A4.5 OR CG11642.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY;		
RX	MEDLINE-20196006; PubMed=10731132;		
RA	Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,		
RA	April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,		
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,		
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,		
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,		
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,		
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,		
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,		
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,		
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,		
RA	Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissensbach J.,		
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,		
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,		
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,		
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;		
RT	"The genome sequence of Drosophila melanogaster."		
RL	Science 287:2185-2195(2000).		
DR	EMBL: AE003419; AAF45569.1; --		
DR	FlyBase: FBgn0040340; EG:BACR7A4.5.		

Db	67	AVFQGEVRHGYLSGILDPLPAIFFYSCWIVHVVQVEYGLDKISKTKHLSKVSFTKFC	126
QY	120	EAQQLSVFVIVSGINGMIIASENCLSD----PTLLWKSP--HNMTTQMFFVYSOL	172
Db	127	ESFH-QMEFTVYSIAHAFIVSER-LEDFSEVKSFSVWLYGTPTEHRVMSAAKLYFIQI	184
QY	173	AYWFHSFPELYFOKVRKODIPQGLIYIGLHFIHGGAVALLYLNHILGLLLMLHVAVELLS	232
Db	185	SYWIHOFPEFYQLKLRDEIQKSVQALHIAFISIAFFNFTRVGLALILEYITQLIF	244
QY	233	SYCSLLYFGDERYQKCLIS-----LWPVFTSGRLVTLIVSVVT--VCLHLACTNRNGNAL	285
Db	245	HIARFAHFGV---RKGSLDPAFKLFNGSFVLRLGSIIAVMTFWYGLRQA-ESPVDVIS	300
QY	286	SGNVNVLAAKIAVLLSSSCSIQWYITWTLLTIVWQLQWLEDANLHVCGRKRSRSGTENG	345
Db	301	AGNFTAVLRNLVLLAVLLOFLYSEVV-----FHM-GRFSENAKREKKKS	348
QY	346	VENPNRIDSPPKKBEK	361
Db	349	AAAAAAV---PKBEKK	361
RESULT	11		
Q9CVJ6			
ID	Q9CVJ6	PRELIMINARY;	PRT; 159 AA.
AC	Q9CVJ6;		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	1810049E02RIK PROTEIN (FRAGMENT).		
GN	1810049E02RIK.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=PANCREAS;		
RX	MEDLINE=21085660; PubMed=11217851;		
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,		
RA	Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,		
RA	Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,		
RA	Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA	Blake J., Boffelli D., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA	Brownstein M.J., Bult C., Fietcher C., Fujita M., Gariboldi M.,		
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,		
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,		
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,		
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,		
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,		
RA	Hayashizaki Y.		
RT	Functional annotation of a full-length mouse cDNA collection.;		
RL	Nature 409:685-690(2001).		
DR	EMBL; AK007839; BAB25296.1; -		
DR	MGD; MGI:1919515; 1810049E02RIK.		
FT	NON_TER		
FT	1		
SQ	SEQUENCE 159 AA; 18053 MW; D88C0B3126B0085A		CRC64;

Query Match 18.4%; Score 348; DB 11; Length 159;
Best Local Similarity 48.1%; Pred. No. 1.4e-23;
Matches 76; Conservative 22; Mismatches 50; Indels 10; Gaps 3;

QY 215 NHGLGILLMLHAYVELLSSVCSLLPYFGDERYQKQLSIPWIFISGRILVTLIIVSVWTGHL 274
|||||:||||||| | : |||:||||||| |: |||:|||||||

Dn 1 NHGLVLLLVHFVFELHISRFYSDEKYQKGFSLMAVLFLVGRLTLILSVLTGVFG 60

OY 275 LA-GTNRNALSGNWNLAAKIAVSSSCSIQTIVITWTLTWLORWLELDANLHVCGRK 333
|| : : : ||| |::|::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Dn 61 LARAENOKLFDSCTGNFNLAVRINVASICTOAFMKMFNFQLRRRHREHSFAFPVPVK 120

OY 334 RR-----SRSRKGTENGVE---NPNRIDSPPKKEKA 362
|:|:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Dn 121 KPAVTTKGRSRRKTENGTVTSNGSDPSPRNKES 158
|:|:|||||:|:|:|:|:|:|:|:|:~::~|:

RESULT 12 PRELIMINARY; PRT; 393 AA.

R9D6J1 ID O9D6J1 AC Q9D6J1 DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DI 01-DIC-2001 (TEMBLrel. 19, Last annotation update)
DE 2900019C14RIK PROTEIN (RIKEN CDNA 2900019C14 GENE) (TRH1).
DN 2900019C14RIK.
OS Mus musculus [Mouse].
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawaj A., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Miyasawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Anon H., Baldarelli R., Barsh G.,
RA Blake J., Borrelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Humé D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ringwald M., Rodriguez J., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawasaki H., Kohtsuki S.,
RA Hayashikuni Y.;
RT Functional annotation of a full-length mouse cDNA collection." ;
RL Nature 409:685-690(2001).
RN [2]
RN SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RA Hartmann E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK013554; BAB28903.1; -
DR EMBL; BC003946; AAA03946.1; -
DR EMBL; AY029531; AAH04299.1; -
DR MGd; MG1.1914510; 2900019C14RIK.
DR InterPro; IPRO01356; Homeobox.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
SQ SEQUENCE 393 AA; 46016 MW; B2038F7A128F816F CRC64;

Query Match	8.6%	Score 163.5	DB 11	Length 393
Best Local Similarity	26.0%	Pred. No. 1.5e-06		
Matches	46	Conservative 40	Mismatches 76	Indels 15
Gaps	5			
QY	106	RLQLTQKQKNLNEAGOLSVFYIVSGIWGMIILASENCLSDPTLLMKSQPHNMFTQMK	165	
Db	121	RNRDPSLSKKKCEACWRFFYVLCSPFGVGTSLYHSRWLWSPSLCWNYPHOTLNLSLY	180	


```
SQ SEQUENCE 400 AA; 46351 MW; AB586DB2B811E62 CRC64;

Query Match 6.3%; Score 119.5; DB 5; Length 400;
Best Local Similarity 18.8%; Pred. No. 0.014;
Matches 49; Conservative 52; Mismatches 96; Indels 63; Gaps 7;

QY 61 AEGLPSSRTLYHVGKDLATVFYMLVATIIHATIOEY----- 99
Db 25 ADIAPGSRPDVHANYKDLIWPIDFAAVMLVRYTLERFWISPVGKSLGIRSRPKKAAN 84
QY 100 --VLDK-LSRRLQLTGKQKNLNEAGOLS-----V 126
Db 85 VPILEKTYAKSTRLDKKKLVPLSKQTDMSEREIERWRLRRAQDKPSTLVKFCENTWRCI 144
QY 127 FYIVSGIWGMIIASENCLSDPTLLWKSQPHNMWTFOMKEFY-ISQLAYWFHSFPPELYFQ 185
Db 145 YLYSFTFGVIVLWDKPWFVDKSCWYGYPHQSISNDIWWYIMISMFYW--SLTGTQFF 202
QY 186 KVRKODIPGOLIYIGLHFLHIGGAYLLYLNHLGILLMLHYAVELLSSVCSLLYFGDERY 245
Db 203 DVRRKDFQMFHMMVTLMLSLSWCNEHRVGSVLVVDHCADIFLEAAKLTKYA--KY 260
QY 246 QK---GLSLWPIVFIISGRL 261
Db 261 QKLCDAIFAIFTVVWIVTRL 280
```

Search completed: September 6, 2002, 17:20:26
Job time: 635 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 14:46:34 ; Search time 5654.61 Seconds
(without alignments)
8467.413 Million cell updates/sec

Title: US-09-807-470-3

Perfect score: 2288

Sequence: 1 tataggcacgcgtggtcga.....ccacttgtaaatgggaattc 2288

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_nu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description

c	1	2254.8	98.5	165538	9	AC093805	AC093805 Homo sapi
2	1288	56.3	1288	6	BD005228	BD005228 Method of	
3	1110	48.5	1110	6	BD005226	BD005226 Method of	
4	961.8	42.0	2856	4	CFTRAM	X63678 C.familiar	
5	937.6	41.0	2722	9	BC000687	BC000687 Homo sapi	
6	804.8	35.2	2387	4	BTU19578	U19578 Bos taurus	
7	803.8	35.1	155623	2	AC106086	AC106086 Rattus no	
8	742.4	32.4	1267	6	AX329795	AX329795 Sequence	
9	742.4	32.4	1267	6	BD005229	BD005229 Method of	
10	742.4	32.4	1267	9	HSTRAMP	X63679 H.sapiens m	
11	733	32.0	2720	10	AY029764	AY029764 Mus muscu	
12	733	32.0	2819	10	BC012401	BC012401 Mus muscu	
13	648.2	28.3	73339	2	AC027246	AC027246 Homo sapi	
14	647.8	28.3	1092	6	BD005227	BD005227 Method of	
15	508.8	22.2	1338	5	AY029534	AY029534 Xenopus l	
16	486.8	21.3	642	6	AX303127	AX303127 Sequence	
17	443.8	19.4	1453	5	AY029529	AY029529 Danio rer	
18	297.8	13.0	6974	9	HUMORFKG1A	D31762 Human mRNA	
19	269	11.8	1360	10	AY029530	AY029530 Mus muscu	
20	269	11.8	1408	10	BC018212	BC018212 Mus muscu	
21	258.4	11.3	1272	5	AY029535	AY029535 Xenopus l	
22	233.4	10.2	196349	2	AC022032	AC022032 Homo sapi	
23	231.8	10.1	163997	9	AC022731	AC022731 Homo sapi	
24	161	7.0	2123	3	AY075393	AY075393 Drosophil	
25	161	7.0	132910	2	AC014319	AC014319 Drosophil	
26	161	7.0	146717	3	AC097726	AC097726 Drosophil	
27	161	7.0	212370	3	DMBR7A4	AL109630 Drosophil	
28	161	7.0	299275	3	AE003419	AE003419 Drosophil	
29	118	5.2	447	6	AX192932	AX192932 Sequence	
30	118	5.2	196349	2	AC022032	AC022032 Homo sapi	
31	94.8	4.1	64530	2	AC110042	AC110042 Mus muscu	
32	93.8	4.1	189572	2	AC084812	AC084812 Homo sapi	
33	88.4	3.9	494	6	AX305631	AX305631 Sequence	
34	87.2	3.8	41805	3	CBRG44C02	AC084610 Caenorhab	
35	81	3.5	420	11	G37645	G37645 SHGC-38238	
36	69.8	3.1	1017	9	AF274933	AF274933 Homo sapi	
37	65.6	2.9	602	6	AX285010	AX285010 Sequence	
38	64.2	2.8	189572	2	AC084812	AC084812 Homo sapi	
39	63.2	2.8	109813	9	HSJ180B22	AL049611 Human DNA	
40	62.2	2.7	14544	3	CEC24F3	AL022716 Caenorhab	
41	60.8	2.7	2468	8	YSCMTORIK	M11449 Yeast mitoc	
42	60.8	2.7	21153	8	YSCMTGCI3	L36897 Saccharomyc	
43	59.4	2.6	155973	2	AC100727	AC100727 Mus muscu	
44	58.2	2.5	7218	6	I66494	I66494 Sequence 14	
45	57.2	2.5	97672	8	AC006951	AC006951 Arabidops	

ALIGNMENTS

RESULT 1

AC093805/c

LOCUS

DEFINITION

AC093805 AC012532

AC093805.3

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

AC093805 Homo sapiens BAC clone RP11-32609 from 4, complete sequence.
165538 bp DNA linear PRI 08-FEB-2002

AC093805 AC012532

AC093805.3

HTG.

human.

human.

human.

human.

human.

human.

human.

human.

human.

human.

human.

human.

Genome Res. 8 (11), 1097-1108 (1998)
Toward a complete human genome sequence
Sulston, J.E. and Waterston, R.
1 (bases 1 to 165538)
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 165538)
2 (bases 1 to 165538)
Cedroni, M., Abbott, A. and Bielicki, L.
The sequence of Homo sapiens BAC clone RP11-32609
Unpublished (2001)
3 (bases 1 to 165538)
Waterston, R.H.
Direct Submission

JOURNAL

Submitted (10-SEP-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

4 (bases 1 to 165538)
Waterston,R.H.

AUTHORS

Direct Submission

JOURNAL

Submitted (05-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

5 (bases 1 to 165538)
Waterston,R.

AUTHORS

Direct Submission

TITLE

Submitted (08-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

JOURNAL

On Feb 5, 2002 this sequence version replaced gi:16077043.

COMMENT

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_NH0326009
Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-32609; actual end is at base position 165538 of RP11-32609.

Data from AC013556 was used to finish the clone, AC093805.

Polymorphisms have been identified between AC013556 and AC093805.

The sequence of AC012332 has been incorporated into AC093805.

FEATURES

source

Location/Qualifiers

1..165538
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-32609"
/clone_lib="RPCI-11"
1..130
/rpt_family="MaLR"

repeat_region

repeat_region 131..436
/rpt_family="ERV1"
repeat_region 870..1245
/rpt_family="L1"
repeat_region 1246..2720
/rpt_family="L1"
repeat_region 2721..3067
/rpt_family="L1"
repeat_region 3627..3713
/rpt_family="MaLR"
repeat_region 4217..5074
/rpt_family="Achoho"
repeat_region 5055..5101
/rpt_family="(TTTA)n"
repeat_region 5075..5384
/rpt_family="Alu"
repeat_region 5385..5592
/rpt_family="Achoho"
repeat_region 7152..7426
/rpt_family="L1"
repeat_region 7427..7734
/rpt_family="Alu"
repeat_region 7590..7619
/rpt_family="AT-rich"
repeat_region 7735..7872
/rpt_family="L1"
repeat_region 7873..8314
/rpt_family="ERV1"
repeat_region 8315..9564
/rpt_family="L1"
repeat_region 9438..9470
/rpt_family="(A)n"
repeat_region 9822..9853
/rpt_family="(CA)n"
repeat_region 10274..10722
/rpt_family="MaLR"
repeat_region 11045..11163
/rpt_family="CT-rich"
repeat_region 13136..13287
/rpt_family="Alu"
repeat_region 15545..15567
/rpt_family="AT-rich"
repeat_region 15850..15875
/rpt_family="AT-rich"
repeat_region 15929..16234
/rpt_family="Alu"
repeat_region 17161..17194
/rpt_family="AT-rich"
repeat_region 18540..18828
/rpt_family="MaLR"
repeat_region 18938..19023
/rpt_family="MaLR"
repeat_region 19181..19224
/rpt_family="MaLR"
repeat_region 19256..19588
/rpt_family="MaLR"
repeat_region 19659..19726
/rpt_family="MaLR"
repeat_region 19847..20312
/rpt_family="MER2_type"
repeat_region 19928..19960
/rpt_family="(TG)n"
repeat_region 20313..20623
/rpt_family="Alu"
repeat_region 20596..20640
/rpt_family="(A)n"
repeat_region 20779..20802
/rpt_family="AT-rich"
repeat_region 20868..21156
/rpt_family="Alu"
repeat_region 21487..21514
/rpt_family="AT-rich"
repeat_region 21958..21983

repeat_region	/rpt_family="AT_rich" 22010..22076	QY	571	tgtttcttcctgctgggctgtgttgcagagggaacagcagagcatccatcgtgtttctca	630
repeat_region	/rpt_family="AT_rich" 22223..22248	Db	48860	TGTTCTTCCCTGCTGGGCTGTGTTCGAGGGAACAGCAGAGAATCCATCGTGTCTCA	48801
repeat_region	/rpt_family="(TTTA)n" 22234..22520	QY	631	ctcttcagacagtggttgctgtccctgcagcagagggaacacagcggtctcaagtccc	690
repeat_region	/rpt_family="Alu" 23084..23385	Db	48800	CTCTTTCAGCAGGTGTGCTGCCCTGCAGCAGAGGAACAGCCACGGGTCTCAAGTCCC	48741
repeat_region	/rpt_family="Alu" 23503..23691	QY	691	tctattattggtgtcaagattggccacggttttcttctacatcgtggtggaatca	750
repeat_region	/rpt_family="CR1" 23715..23895	Db	48740	TCATATTATTATGGTGTCAAGATTTGGCCACGGTTTCTTCTACATGCTGGTGGCAATCA	48681
repeat_region	/rpt_family="MIR" 24054..24855	QY	751	ttattcatgcacaattccaggaatatgtttggataaaaaataacaagagaatgcagttca	810
repeat_region	/rpt_family="CR1" 24966..25190	Db	48680	TTATTTCATGACCAATTCAGGAATATGTTGGGATAAAATTAACAAGAGAATGCAGTTCA	48621
repeat_region	/rpt_family="L2" 25636..25670	QY	811	ccaaagcgaaacaaacaagttaaocagtgctggtcagtttagtggttcttactttttt	870
repeat_region	/rpt_family="(CA)n" 25704..26073	Db	48620	CCAAAGCGAAACAAAACAAGTTTAAACGAGTGTGCTCAGTTTGTCTTACTTTTTT	48561
repeat_region	/rpt_family="MaLR" 26711..27112	QY	871	ctgtatttggggcacattcatcttaatactctgaaaaactcgcctgcagacccaactcta	930
repeat_region	/rpt_family="MaLR" 27461..27664	Db	48560	CTTGTATTGGGGCACATTCAATTTAATCTCTGAAAACCTGCCTGTACAGCCCAACTCTTA	48501
repeat_region	/rpt_family="L2" 28756..28814	QY	931	tatggaaggctcgtccccatagcatgatgacattccaattcaaatgaagtctttctacataccc	990
repeat_region	/rpt_family="L2" 29356..29574	Db	48500	TATGGAAGGCTCGTCCCCATAGCATGATGACATTTCAATGAAAGTTTTTCTACATATCCC	48441
repeat_region	/rpt_family="L1" 29748..29876	QY	991	agttggttactggtttctcgtcttctcgaactctacttccagagaaacccaaaaacaag	1050
repeat_region	/rpt_family="ERVL" 30292..30476	Db	48440	AGTTGGCTTACTGTTTTCATGCTTTTCTGAACTCTACTTCCAGAAAAACAAAAACAAG	48381
repeat_region	/rpt_family="L1" 98.58; Score 2254.8; DB 9; Length 165538; Best Local Similarity 99.9%; Pred. No. 0; Matches 2256; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	QY	1051	acatccctctcaacttgctacatgtgtcttcacctcttccacattactggagcttctc	1110
		Db	48380	ACATCCCTCGTCAACTTGTCTACATTTGGTCTTCACCTCTTCCACATTACTGGAGCTTATC	48321
		QY	1111	tcttgacttgaaatcatttgggactctctcttgggtactgcattatttggtagaattac	1170
		Db	48320	TCATTGTACTTGAATCATTTGGGACTTCTCTTTTGGTACTGTCATTATTGTTGAATTAC	48261
		QY	1171	tttccacatgtcgggcctgttttactttagtgatgaaaagtaccagaaaggcatactc	1230
		Db	48260	TTTCCACATGTGCGGCTGTTTACTTTAGTGATGAAAAGTACCAGAAAAGCATATCTC	48201
		QY	1231	tgtgggcattgtttatctgtgggtagacttgcactttaattgttccgtactcactg	1290
		Db	48200	TGTGGGCCATTGTGTTTATCTTGGGTAGACTTGTGACTTTAAATGTTTCCGTACTACTG	48141
		QY	1291	ttgggtttccacctggctggatcgagaaatcggaatcctgatgccttacttggaaatgaa	1350
		Db	48140	TTGGTTFACCTGGCTGGATCGCAGAATCGGAATCCTGATGCCCTTACTGGAATCTAA	48081
		QY	1351	atgtgttggcagctaaattgcgttctgtcgtccagttgcagatccagcgtcaatgaa	1410
		Db	48080	ATGTGTTGGCAGCTTAAATTTGCTTCTGTCGTCCAGTTGCACGATCCAAAGCCTACGTAA	48021
		QY	1411	catggaacttaatactctctgcaggttcagaggtgggttagagattctaattatcagccct	1470
		Db	48020	CATGGAACCTTAAATTACTTCTTGGCTTCAGAGGTGGGTAGGAAGATTCTTAATTACAGCCCT	47961
		QY	1471	catgtagaaaaaacggtcgagatcttcttaaaaaaacacagaaacccgagtgaggag	1530
		Db	47960	CATGTATGAAAAGAAACGCTCGAGATCTTCTTAAAAAAGACACAGAAACCGAGTGGGAG	47901
		QY	1531	tggaaacttcaaatagatgtagactgtccgcgaagaggaagagaaatcttctaatacttt	1590
		Db	47900	TGAAAACCTTCAAAATAGAGTAGACTGTCCGCCAAAAGAGAAAGAGAAATCTTCATAATCTT	47841
		QY	1591	tgaaggcattgatttaattgtctgcaaaaggaatctgctcttttgaggtttctttctgcaat	1650
		Db	47840	TGCAAGCGCATTGATTAAATGCTCTCAAAAGGAATCTGCTCTTTGAGGTTCTTTCTTGCACT	47781

Qy	1651	agagatttttctgttttgaaatagttcgtctcttcggtttttgtttattgaactgttt	1710
Db	47780	AGAGATTTTCTGTTTGTGAAATAGTTCGTCTTCGGTCTTTTGTATTGAACGTGT	47721
Qy	1711	catgtatttttaagacatttgagggaagagattattatgaatgggaaaaagatt	1770
Db	47720	CATGTATTTTAAAGACATTTGAGGGAGGAGGATTATTATGAATGGGAAAAAGATT	47661
Qy	1771	ttggtgagactaaattactcatcgtcaaaataatgccaataagtttttgggacacca	1830
Db	47660	TTGGTTGAGACTAAATPACTCATCGTCAAAATAATGTCAAAATAGTTTGGGATCACCA	47601
Qy	1831	ctatatttgttttgatttttaacctttcaacattttccctaatttcagagataact	1890
Db	47600	CTATATTTGTTTGTATTTAACCTTCAACATTTTCCCTAATGATTTGCAGAGATAACT	47541
Qy	1891	gcacaatttgcatatacaatgatactgggttcttactccccaccaggttttcataactaa	1950
Db	47540	GCACAATTTTGCAATPATAATGATACTGGTCTTACTCTCCACCACTGTTTCATAATACTAA	47481
Qy	1951	caagatggtctctctagcaagattatgttttaagtcttgctttgggtaaaaaag	2010
Db	47480	CAAGATGGTCTCTAGCAAGATTATGTGTTAATGCTTGTGTTGGGGTAAAAATAAG	47421
Qy	2011	tacgaaaagtggaagtcacaaatcagtattctgttaattgttagaatttatttttaagaa	2070
Db	47420	TACGAAAAGGTGGAAGTCAAAATCAGTATCTGTAATGTTAGATTTATTTTAAAGAA	47361
Qy	2071	cttaactcagaaaaagattgtagactcaccacaaataataaatgttctttatttcacag	2130
Db	47360	CTTACAACCTCAGAAAAGATTGCTAGACTCACCAAAATAATAAATGTTCTTTATTTACAG	47301
Qy	2131	gtagtgattattagtggttcattcccatcttaaaaaaacacagttactaatgggtaacacat	2190
Db	47300	GTAGTGATTATTAGTGCTTCATCCCCATTTAAAAAACACAGTACTAATGGGTAAACAT	47241
Qy	2191	atggaggttgctgccatatatttgatcaaaataatcatttaataataaaaaatatta	2250
Db	47240	ATGGAGGTTTGCTGCCATATATTGTCATCAAAATATCATTAATATATAAANAATTA	47181
Qy	2251	aaatcatctcgtccatccactgttaaatggggaattc	2288
Db	47180	AAATCATCTGTGTCATTCACCTTGTAATGGGAATTC	47143
RESULT 2			
LOCUS	BD005228	1288 bp	DNA linear PAT 31-JAN-2002
DEFINITION	Method of screening of protein.		
ACCESSION	BD005228		
VERSION	BD005228.1	GI:18633189	
KEYWORDS	JP 03075332-T/3.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 1288)		
TITLE	Todo,N., Okuyama,H., Imamura,M., Ishikawa,H. and Nemoto,K.		
JOURNAL	Method of screening of protein		
COMMENT	Patent: JP 03075332-T 3 16-FEB-2001;		
	SUMITOMO PHARMACEUTICALS CO LTD, NAOKI TODO, HAJIME OKUYAMA, OTOAKI		
	IMAMURA, HIROKORI ISHIKAWA, KIYOMITSU NEMOTO		
	OS Homo sapiens (human)		
	PN JP 03075332-T/3		
	PD 16-FEB-2001		
	PF 17-AUG-2000 JP 2000005488		
	PR 20-AUG-1999 JP 99P 234764		
	PI NAOKI TODO, HAJIME OKUYAMA, MOTOAKI IMAMURA, HIROKORI ISHIKAWA,		
	KIYOMITSU NEMOTO		
	PC C1201/02.G01N33/50.C07K14/47.A61K38/17.C12N5/10.C12P21/02// PC		
	(C12P21/02,C12R1:91)		
CC			
PH	Key	Location/Qualifiers	

FEATURES	FT	CDS	Location/Qualifiers	(40)..(1146).
source			1..1288	
			/organism="Homo sapiens"	
			/db_xref="taxon:9606"	
BASE COUNT	331 a	275 c	276 g	406 t
ORIGIN				
			Query Match	56.3%; Score 1288; DB 6; Length 1288;
			Best Local Similarity	100.0%; Pred. No. 1.1e-300;
			Matches 1288; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	438	caactcgccgctcgcgcgtgagttatccgggaagcgcaccatggggctccgtaagaagacg	497	
Db	1	CAGCTGCGGCTCGCGCGTGAGTATCCGGGAAGCGCCACCATGGGGCTCCGTAAAGAGAC	60	
Qy	498	accaaagaaccccccggttctcagccaggaaattcatctcgagaatcatcgggacatcgtc	557	
Db	61	ACCAAGAACCCTCCGCTTCTCAGCCAGGAATTTCATCTGCAGAAATCATGCGGACATCGTC	120	
Qy	558	tcctcgctggggatgttcttctcgtctggggcttgtgttcagagggaaacagcagaagcatcc	617	
Db	121	TCCTCGCTGGGATGTCTTCCTGCTGGGGCTTGTGTTTCGAGGGAACAGCAGAAAGCATCC	180	
Qy	618	atcgtgttctcactcttcagcacagttgtctgtccctcagcagaggaacaaagccacg	677	
Db	181	ATCGTGTTCCTCACTTCTCAGCACAGTGTGCTGTCCCTGCAGCAGAGGAACAAAGCCACG	240	
Qy	678	ggctcaaaagtcctctctattattatggtgtcaaaagatttggccacggtttcttctcacatg	737	
Db	241	GGCTCAAAGTCCCTCTATTATTATGTTGTCAAAGATTGGCCACGGTTTCTTCTACATG	300	
Qy	738	ctggtggcaatcatcattatcctgcacaattccaggaatatgtgtggataaaattcaacaag	797	
Db	301	CTGCTGGCAATCATATTTCATGCCACAATTTCAGGAATATGTGTGGATAAATAAACAAG	360	
Qy	798	agaatcagttcccaagcgaacaaacaaagtttaacagagtcgtcagtttagttgtg	857	
Db	361	AGAAATGCAGTTTACCAAGCGGAACAAACAAAGTTTAAACAGTCTGGTCAAGTTAGTGTG	420	
Qy	858	ttctactttttctgttatttggggcacattctttaaactctctgaaaactgcctgtca	917	
Db	421	TTCTACTTTTTTCTGTGATTTTGGGCACATTTCATTTAATCTCTGAAAACCTGCTGTCA	480	
Qy	918	gaccacactcttataggaaagctcgtcccatagcatgatgcatttcaaatgaagtgt	977	
Db	481	GACCCCAACTCTTATATGGAAGGCTCGTCCCATAGCATGATGACATTTCAAAATGAAGTTT	540	
Qy	978	ttctacatatccagttgcttactggttttcattgcttttctgaaacttacttccagaaa	1037	
Db	541	TTCTACATATCCAGTTGGCTTACTTGGTTTTCATGCTTTTCTGAACTCTACTTCCAGAAA	600	
Qy	1038	accaaaaacaagacatccctcgtcaactgtctacatttggtcttcaactcttccacatt	1097	
Db	601	ACCAAAAAACAAGACATCCCTCGTCAACTTGTCTACATTTGGTCTTCACCTCTTCCACATT	660	
Qy	1098	actggagcttatctctgtacttgaatcatttgggactcttcttcttggtagctattat	1157	
Db	661	ACTGGAGCTTATCTCTGTACTTGAATCATTTGGGACTTCTTCTTTTGGTACTGCATTAT	720	
Qy	1158	tttgtgaattactttccacatgtcgccgctgttttacttttagttgatgaaaagaccag	1217	
Db	721	TTTGTGTGAATTACTTTTCCACATGTGGGGCTGTTTTTACTTTTGTGATGAAAAGTACCAG	780	
Qy	1218	aaaggcatatctctgtgggccatttgttttatctctgggtagacttgtgactttaattgtt	1277	
Db	781	AAAGGCATATCTCTGTGGGCCAATGTGTTTATCTTGGTAGACTTGTGACTTTTAATTGTT	840	
Qy	1278	tcogtactcactgttgggttttcaactggctggatcgcagaatcgggaatcctgatgcctt	1337	
Db	841	TCCGTACTCAGTGTGGGTTTTCAGCTGGCTGGATCGGGAATCGGAATCTGTATGCCCTT	900	


```
QY 1338 actggaatgtaaatgtgttgagcagctaaattgctgttctgtcgtccagtgtaacagatc 1397
Db 901 actggaatgtaaatgtgttgagcagctaaattgctgttctgtcgtccagtgtaacagatc 960
QY 1398 caagcctaagtaacatggaacttaatactctctgtggtcgaaggtggttagaagattct 1457
Db 961 caagcctaagtaacatggaacttaatactctctgtggtcgaaggtggttagaagattct 1020
QY 1458 aatattcagcctcatgtatgaaagaaacaggtcagatcttctaaataaaagaaacagaa 1517
Db 1021 aatattcagcctcatgtagtaaaagaaacaggtcagatcttctaaataaaagaaacagaa 1080
QY 1518 aacggagtggagtggaacttcaaatagatgtagctgcccgaagaggaagagaa 1577
Db 1081 aacggagtggagtggaacttcaaatagatgtagctgcccgaagaggaagagaa 1140
QY 1578 tcttcataatctttgcaagcgaattgatttaattctgcaaggaatctgctctttgaggt 1637
Db 1141 tcttcataatctttgcaagcgaattgatttaattctgcaaggaatctgctctttgaggt 1200
QY 1638 tctttctgactagagattttctgttttgaataatgctgctcttctggttttgg 1697
Db 1201 tctttctgactagagattttctgttttgaataatgctgctcttctggttttgg 1260
QY 1698 tattgaactgtttcatgtatttttaa 1725
Db 1261 tattgaactgtttcatgtatttttaa 1288

RESULT 3
BD005226 1110 bp DNA linear PAT 31-JAN-2002
LOCUS Method of screening of protein.
DEFINITION BD005226
ACCESSION BD005226
VERSION BD005226.1 GI:18633187
KEYWORDS JP 03075332-r/1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1110)
Todo,N., Okuyama,H., Imamura,M., Ishikawa,H. and Nemoto,K.
Method of screening of protein
Patent: JP 03075332-T 1 16-FEB-2001;
SUMITOMO PHARMACEUTICALS CO LTD, NAOKI TODO, HAJIME OKUYAMA, OTOAKI
IMAMURA, HIRONORI ISHIKAWA, KIYOMITSU NEMOTO
OS Homo sapiens (human)
PN JP 03075332-r/1
PD 16-FEB-2001
PF 17-AUG-2000 JP 2000005488
PR 20-AUG-1999 JP 99P 234764
PI NAOKI TODO, HAJIME OKUYAMA, MOTOAKI IMAMURA, HIRONORI ISHIKAWA,
PI KIYOMITSU NEMOTO
PC C12Q1/02.G01N33/50.C07K14/47.A61K38/17.C12N5/10.C12P21/02// PC
(C12P21/02.C12R1:91)
CC
FH Key Location/Qualifiers
FT CDS (1)..(1107).
FEATURES
source Location/Qualifiers
1..1110
/db_xref="taxon:9606"
BASE COUNT 296 a 241 c 236 g 337 t
ORIGIN

Query Match 48.5%; Score 1110; DB 6; Length 1110;
Best Local Similarity 100.0%; Pred. No. 1.2e-257;
Matches 1110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 atggggctccgtaagagcaccagaaccccccttctcagccaggaattcctg 536
Db 1 atggggctccgtaagagcaccagaaccccccttctcagccaggaattcctg 60
```

```
QY 537 cagaatcagcgacatcgtctcctcgtggggatgttctctcgtcgtgggctgtgttc 596
Db 61 CAGAAATCATCGGACATCGTCTCCTCGTGGGATGTTCTTCTGCTGGGCTGTGTTC 120
QY 597 gagggaaacagcagaagcaccatccatcgtgttctcactcttcagcacagtggtgtgcct 656
Db 121 GAGGGAACAGCAGAGAATCCATCGTGTCTTCTCAGCAGAGTGTGTGCTGCCCT 180
QY 657 gcagcagaagaacacagcggtcgaagtcctctattattatggtgcaagatttg 716
Db 181 GCAGCAGAGAAACAGCCAGCGGCTCAAGTCCCTCTATTATTGCTGTCAAGATTTG 240
QY 717 gccacggtttctctacatgctgtggcaatcatctattcatcccaatcaggaaat 776
Db 241 GCCACGGTTCCTCTACATGCTGGTGGCAATCATATTATCATGCCACAATTCAGGAATAT 300
QY 777 gttgtgataaaattacaagagaaatgcagttccacaaagcgaacaaacaaagttaac 836
Db 301 GTGTTGGATAAAATTAACAAGAGAATGCAGTTCCACCAAGCGAAACAAAAGTTTAAAC 360
QY 837 gagctcgtcagtttagtggttctacttttttttttttttttttttttttttttttt 896
Db 361 GAGTCTGGGTGAGTTAGTGTGTCTTCTTTTCTTGTATTGTTGGGCACATTCATTTTA 420
QY 897 atctctgaaactgcctgtcagaccacactcttatatggaagctcgtcccatagcatg 956
Db 421 ATCTCTGAAACATGCTGTGAGACCCCACTTATATGGAAGGCTGCTGCCCATAGCATG 480
QY 957 atgacatttcaaatgaagttttctacatatccagttggtcgttactgttctatgctttt 1016
Db 481 ATGACATTTCAAAATGAAGTTTCTACATATCCAGTTGGCTTACTGGTTTCATGCTTTT 540
QY 1017 cctgaactctctcagaaacacaaacacacacacacacacacacacacacacacacacac 1076
Db 541 CCTGAACCTCTACTCTCAGAAACCAAAACAAAGACATCCCTCGTCAACTGTCTACATT 600
QY 1077 ggtcttcacactctcacattactgaggttatctctgttacttgatcatcttttgggactt 1136
Db 601 GGTCTTCACTCTCTCACATTTACTGGAGCTTATCTCTGTACTTGAATCATTTGGGACTT 660
QY 1137 cttcttttggtactgattttttgttgaaataactttccacatgctggtcgtgttttac 1196
Db 661 CTTCTTTTGGTACTGCAATTTTGTGTGAATTTCTTCCACATGTGCGGCCCTGTTTAC 720
QY 1197 ttttagtgaaagaccagaaagcagatctctctgtggcgaattgtgtttatcttggtt 1256
Db 721 TTTTAGTGATGAAAGATACCAGAAAGGCATATCTCTGTGGGCCATTGTGTATTATCTTGGGT 780
QY 1257 agactgtgactttaaattgtttccgtactactcgttgggtttccacctggctggatcgag 1316
Db 781 AGACTTGTGACTTTAATTGTTCCGTACTACTCTGTTGGGTTTCACTGGCTGGATCGCAG 840
QY 1317 aatcggaactcctgatgcccttactggaataatgtaaatgtgttggcagcgttaaatgtgtt 1376
Db 841 AATCGGAATCCTGATGCCCTTACTTGGAAATGTAAATGTGTGTGGCAGTAAATTTGCTGT 900
QY 1377 ctgtcgtccagttgacgatacgaagcctacgttaaacatggaacttaattactctcgtgtt 1436
Db 901 CTGTCTCCAGTTGCAGATCCAGGCTACGTAAACATGGAACCTTAATTAATCTCTGCTT 960
QY 1437 cagagtggttagaagattcttaattcagggcctcatgtatgaaagaaacagcgttcgaga 1496
Db 961 CAGAGTGGGTAGAAATCTTAATATTCAGGCTCATGTATGAAAGAAAGAAACGGTCGAGA 1020
QY 1497 tcttctaaaaaagacagaaacggagtggtggagtggaacttcaaatagtagactgt 1556
Db 1021 TCTTCTAAAAAAGAACAGAAACGGAGTGGGAGTGGAAACTTCAAATAGAGTAGACTGT 1080
QY 1557 cogccaaagaggaaagagaaattctcataa 1586
Db 1081 CGCCCAAGAGGAAAGAGAAATCTTCATAA 1110
```

```

RESULT 4
CFTRAM      2856 bp      mRNA      linear      MAM 02-JUN-1992
LOCUS       C.familiaris mRNA for TRAM-protein.
DEFINITION  X63678
ACCESSION   X63678
VERSION     X63678.1 GI:941
KEYWORDS    TRAM-protein.
SOURCE      dog.
ORGANISM    Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1 (bases 1 to 2856)
AUTHORS    Gorlich,D., Hartmann,E., Prehn,S. and Rapoport,T.A.
TITLE       A protein of the endoplasmic reticulum involved early in
            polypeptide translocation
JOURNAL     Nature 357 (6373), 47-52 (1992)
MEDLINE     92244357
REFERENCE   2 (bases 1 to 2856)
AUTHORS    Hartmann,E.
TITLE       Direct Submission
JOURNAL     Submitted (28-JAN-1992) E. Hartmann, Max-Delbrueck-Centr. f.
            Molekulare Med., Robert-Rössle-Strasse 10, O-1115 Berlin Buch, FRG
FEATURES   Location/Qualifiers
            1..2856
             /organism="Canis familiaris"
             /db_xref="taxon:9615"
             /cell_line="MDCK"
             219..1343
              /codon_start=1
              /product="TRAM-protein"
              /protein_id="CA445217.1"
              /db_xref="GI:942"
              /db_xref="SWISS-PROT:Q01685"
              /translation="MAIRKSKTSPEVLVSHEFTLQNHADIIVSCVMVFLGLMFEITA
                KASIIIFQYNTWLPATEQATPESTLVYYGKDLATVFFYMLVAIIHAIIOEVL
                DLKNRRHESKTKHSFNEGSGLSAYFLSCWGTFLILSENVISDPTILWRAYPHN
                MTFQMFVEYIAQLWFAHPPELYFOKTKKEDIPLQVLVYIGLYLPHIAGAYLLNLNLH
                GLVLLVHLVEFLHLSRLFYSDEKYGKSLMAVLFVLGRLLTLILUSLVTVGFGFL
                ARAENKLFDSAGNFVNLAVIAVLASICITQAFMMKVFINFQLRWRHESHTFOAPWV
                KKAPTIVTKGRSHKKGTEGVNGVTGTSNADSPNRKES"
BASE COUNT 830 a 517 c 532 g 977 t
ORIGIN

Query Match      42.0%; Score 961.8; DB 4; Length 2856;
Best Local Similarity 73.2%; Pred. No. 1.1e-221;
Matches 1447; Conservative 0; Mismatches 462; Indels 68; Gaps 14;

QY 350 gcgaccgcggactgacttcgggtagggaagtggagccgccccggagctgctaccctggcggc 409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92 GCGGACGCGGGCGGTGAGGAGCAGCCAGGGGAGGCGGCGCGCGCGCGCGCGCGCAGC 151

QY 410 ggcgcgtgtgaggagcgaccagggggaggcagctgcgcgtgcgcgtgtagtatccgggaaag 469
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 CAGTCTGTGAGCAGCTGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 211

QY 470 cgcacaccatgggctcgttaagaagagcaccagaagaccccccggttctcagccaggaatt 529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212 CCTCGCATGGCGATTGCGAAGAGAGACCAACAGACCCCCCGCGGTCTGAGCCACCGAAATT 271

QY 530 catctgcagaatcatcgcgacatcgtctctcgtcgctggggatgttctctcgtcggggct 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 272 CATCTCGCAAAATCAGCGGACATCGTCTCTGTGTGGCGATGGTCTTCTGCTGGGGCT 331

QY 590 tfgtttcgagggaacagcaagacatccatcgtgtttctcaactctctcagcacagtggttc 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 332 CATGTTGAGATAACGGAAGACCCCTATCATTTTTGTACTCTTCAGTCAAAATGTTAC 391

QY 650 tgtccctgcagcagaggaacagccagggctcaaaagtcctctattattatgggtgca 709
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 392 CCTCCCTGCAACAGAAGAACAGCTACTGAATCAACATCCCTTATTACTATGGTATCAA 451

```



```
Qy 1005 ttccatgcttttccctgaacttacttccagaaaaacccaaaaacagacatccctcgtcaa 1064
Db 620 CTTCATGCTTTTCTGGAAGACTTCTGACTTCCAGAAAACCAAAAAGAGATATTCTCGTCAG 679
Qy 1065 ottgctcatagtgtcttccactctccacattactggaggttatctcttgaattgaat 1124
Db 680 CTGTGCTACATGTGCTTACCTCTCCACATTTGCTGGAGCTTACCTTTTGAACCTGAAT 739
Qy 1125 catttgggaactctctcttggtagctgattatttggtagattacttctccacatgtgc 1184
Db 740 CATCTAGACATGTGCTTCTGGTGCATATTATTGTTGAATTTCTTTCCACATTTCC 799
Qy 1185 ggcctgttttacttttagtgatgaagaagaccagaaagcatatctctgtggccattgtg 1244
Db 800 CGCCTGTTTTATTAGCAATGAAAGATATCAGAAAGGATTTTCTCTGGGCGAGTCTT 859
Qy 1245 tttactctggtagactgtgaactttaattgtttccgtactcoactgttgggtttcoactcg 1304
Db 860 TTTGTTTTGGGAAGACTTCTGACTTTAATCTTTTCAGTACTGACTGTGTGGTTTGGCCTT 919
Qy 1305 gctgatacgcagaatcgaatcctgagccttactggaatgtaaatgtgtgtgagcgt 1364
Db 920 GCAGAGCAGAAATCAGAACCTGGAATTCAGTACTGGAATCTCAATGTGTTAGCTGTT 979
Qy 1365 aaaattgctgtctgtcgtccagttgcaacgacatccaaagcctacgtaaacttaatt 1424
Db 980 AGAATCGCTGTTCTGGCATCCATTTGCGTTACTCAGGCAATTTATGATGGGAAGTTCATT 1039
Qy 1425 actctcgtgcttcagagtggtgagaagattctctaattctcgggctcctaatgataaaaaag 1484
Db 1040 AATTTCAGCTTCAAGAGTGAGGGAACATTTCTGCTTTTCAGGACCACAGCTGTGAAGAAG 1099
Qy 1485 aaacggtcg-----agattctctaaaaaagacagaaaaacggagtg---gga 1529
Db 1100 AAACCAACAGTAACTAAGGCGAGATCTCTAAAAAAGGAACAGAAATGTTGTAATGGA 1159
Qy 1530 gtggaacttcaaatagatgactgtccgcgaagagggaagaaatcttcataat-c 1588
Db 1160 ACATTAATCTCAAAATGTAGCAGACTCTCCCGGAATAAAAAAGAGAAATCTTCATAATGA 1219
Qy 1589 ttgcaagcagattgattaatgtctgcaagaagaaatcgtctcttggaggttcttcttgca 1648
Db 1220 AATTAAACTAATTGATTAATGTCCCAAAAGAAATCTGCTTTCTACTATATCTTTTCAGCA 1279
Qy 1649 ctagagattttctgtttttgaaaaat--agttcgtgctcttcggttttggttattgaact 1706
Db 1280 TTAGAGATTTTCTGCTTCTGAAATACAGTCTGCTCTTTGATTTTGTCTATTGTGACG 1339
Qy 1707 gttcctgattttttt--aaagacatttgaggaggaggtattattatgaatgggaaaaa 1764
Db 1340 GTTTCATGCATTTTAAAGGGCATTTGAGGGAGGATTTATGCTATGAAT--GAAAAA 1397
Qy 1765 aagatttgggttgagactaaattactcactcgcgaataatgtcaaaatagttttgggga 1824
Db 1398 AATATTTTAGCTTAGACTAAGTACCTCGCTTCAAAATA-----GTTTAGGGA 1445
Qy 1825 tcaccactatatttggttgattttaaacctttcaacattttcctaagtatttgcagag 1884
Db 1446 CCACCACCAATATTATTATTG- -TTTTATTATTTTGAACATTTTCTAATGATTTGGAGAG 1503
Qy 1885 ataactgc-----acaatttggcatataatgatactggttcttactcccaccagtggtt 1939
Db 1504 AAAAATAATTACAAAAATTCACATATACAGTGATACAATTTCTTGTCTCACCAATTTT 1563
Qy 1940 cataactaacaagatggtctctccctcagcaagattattggttttaagtctgtttt-ggg 1998
Db 1564 TATAA---TAGCAGAGTGGCCTGTCTTAAGAAAGGCCATATTTTTTAAGTATATCTTTCCAGG 1620
Qy 1999 gtaaaataaaagtcagaaaaaggtggaagtcaaa-----tcagtatctgta 2045
Db 1621 GTACATGGAATACTATAAAGTTGGATGTCAAACTTAATATGTTTTTCAGTGTCTCTA 1680
Qy 2046 atgttgaattttatttttaagaacttacaactcagaaaaagattgttagactcaccaaa 2105
```

```
Db 1681 ATTTTTCGGAATT-TTGTAGACTTTACACCTGGAAAAAAGATTGTGAAATCACCAGA 1739
Qy 2106 ataataaagtgtcttatttttacaggtagtagtattattagttcctcatcccatttt--aaa 2163
Db 1740 ACAAT-TGTGTGCTTTATTTTATAGGTAGTGTGTTATTAGTATTACATCCCATTTTAAAA 1798
Qy 2164 aaacacagtagtaactgggtaacacatagaggttttgcctccatataattgcatcaaa 2223
Db 1799 ACAAAAACATTAATGTTTACACACGCTGGAGTTTACTTAACATACATATTAATAATCAA 1858
Qy 2224 atatcat-taataataaaaaatattaaaaatcaatcctctcctcatt 2268
Db 1859 GTATATTCTTAAAAAGTACTTGTGAAGTAAATCTTTCTGTGTGCATT 1904

RESULT 6
BTU19578 2387 bp mRNA linear MAM 06-SEP-2000
LOCUS Bos taurus translocation chain-membrane associating protein (tram)
DEFINITION mRNA, partial cds.
ACCESSION U19578
VERSION U19578.1 GI:9971727
KEYWORDS
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 2387)
AUTHORS Clark,T.G., Morris,J., Akamatsu,M., McGraw,R.A. and Ivarie,R.D.
TITLE Cloning and sequence analysis of a bovine tram cDNA
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 2387)
TITLE Direct Submission
JOURNAL Submitted (06-JAN-1995) Genetics, University of Georgia, Athens, GA
30602-7223, USA
FEATURES
source 1..2387
/organism="Bos taurus"
/db_xref="taxon:9913"
<1..2387
/gene="tram"
<1..1077
/gene="tram"
/codon_start=1
/product="translocation chain-membrane associating
protein"
/protein_id="AAG10391.1"
/db_xref="GI:9971728"
/translaton="EFVLONHADIVSCVAMVLLGLMFEITAKVSIIFVTLOYNTLP
ATPEQATESAFLYYGIKDKLATVFFYMLVAIIHAIIOEYVLDKINRMHFSKTKHK
FNESGQLSAFLFSCWIGTFLISENYITLWRAYPHNLMTFQMKFYISOLAYW
FHAFPELYFOKTKEDIPQOLVYIGLYFHAGYLLNLHLGLVLLVLYFVFLPH
ISRLFFYDEKYOKQSLWAVLVGLRLLTLLSVLVGFLAENOKLDFSTGFN
VLAVRIATLASICITQAFMMKFTINFOLRRWRHSAFQAPAVKKKPPVTKGRSXXKT
ENGVCNTVTSNGADSPNRKESK"
BASE COUNT 675 a 473 c 449 g 788 t 2 others
ORIGIN
Query Match 35.2%; Score 804.8; DB 4; Length 2387;
Best Local Similarity 76.6%; Pred. No. 9.7e-184;
Matches 1137; Conservative 0; Mismatches 309; Indels 38; Gaps 11;
Qy 525 gaattcctctgcagaatcatcgagacatcgtctcctcgtcggtgggagtggttctcctgctg 584
Db 1 GAATTCGCTCGCAGATCAGCGGACATCGTCTCCTGCTGTGGGATGGTCTTCTCTGCTG 60
Qy 585 gggcttgctgcagagggaacagcagcagatcccatcgtgtttctcactcttcagcacagt 644
Db 61 GGGCTCATGTTCGAGATAACGAAAAAGTTTCTATCATTTTTTGTACTCTTTCAGTACAAT 120
```

Qy	1702	gaactgtttcaigtatttttttaaa--gacatttgaggggagagattattattatgaatggg	1759
Db	1201	GTATTATTTCATGCATTTTTTAAAAAGGCGATTTTGGGGGAGGATGATCCCATGA--GTG	1258
Qy	1760	aaaaaagatttggttgagactaaactactcatcgtaaaaataatgccaataatgatttt	1819
Db	1259	AAAAAATGTTTTTAGCTTAGCTACCTGCTTCAAAATAGTGTAGGAGCACCAT	1318
Qy	1820	gggagaccactactatttggtttggatttttaacctttcaacattttccctaataatgattg	1879
Db	1319	--CATATTATATTTTATTATTATTGTTTTTTTATCTTTAAGCATTTTCTTAATGATTTG	1375
Qy	1880	caggagataact-----gcacaatttgcatacataatgat-actggttcttactccaccaca	1933
Db	1376	AAGAGATAACATTTATTACAAAAATATATACATATCATGATGAATAAATTTCTGCTCTCACCA	1435
Qy	1934	gtgtttcataatactaacaagatggtctctcctacgaagattat	1977
Db	1436	ATTTTATATATTA--AGCAAGATGCCCTGTTCCACGAGGGTCAT	1477
RESULT	7		
AC106086			
LOCUS	AC106086	155623 bp	DNA linear
DEFINITION	Rattus norvegicus clone CH230-126G18, *** SEQUENCING IN PROGRESS		HTG 12-JAN-2002
ACCESSION	AC106086		
VERSION	AC106086.1	GI:18138601	
KEYWORDS	HTG; HTGS_PHASE1.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
AUTHORS	1 (bases 1 to 155623)		
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhaq,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Duggan,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dunn-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loughes,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Monabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenko,S., Ogulu,M., Okwunu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N., Slonson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlecyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,		

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 155623)
Worley, K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHHR
Center clone name: CH230-126G18
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 127470 bases at least Q40
Consensus quality: 134419 bases at least Q30
Consensus quality: 139475 bases at least Q20
Estimated insert size: 122947; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 5824: contig of 5824 bp in length
5825: gap of unknown length
5925: contig of 4369 bp in length
10293: gap of unknown length
10294: contig of 3535 bp in length
10394: gap of unknown length
13929: contig of 5410 bp in length
14029: gap of unknown length
19439: contig of 3187 bp in length
19539: gap of unknown length
22726: contig of 2118 bp in length
22826: gap of unknown length
24943: contig of 4173 bp in length
24944: gap of unknown length
25043: contig of 3699 bp in length
25044: gap of unknown length
29217: contig of 4602 bp in length
29316: gap of unknown length
29317: contig of 2458 bp in length
33016: contig of 2025 bp in length
33115: gap of unknown length
33116: contig of 4602 bp in length
37817: gap of unknown length
37818: contig of 2458 bp in length
40275: gap of unknown length
40276: contig of 2025 bp in length
40376: gap of unknown length
42401: contig of 2523 bp in length
42500: gap of unknown length
45023: contig of 2993 bp in length
45123: gap of unknown length
45124: contig of 3513 bp in length
48117: gap of unknown length
48217: contig of 2760 bp in length
51729: gap of unknown length
51823: contig of 2843 bp in length
54589: gap of unknown length
54590: contig of 3065 bp in length
57532: gap of unknown length
57533: contig of 1821 bp in length
57633: gap of unknown length
60697: gap of unknown length
60698: contig of 1078 bp in length
62618: gap of unknown length
62719: contig of 2433 bp in length
65151: gap of unknown length
65152: contig of 2592 bp in length
67843: gap of unknown length
67943: contig of 1321 bp in length
67944: gap of unknown length
69264: contig of 2483 bp in length
69265: gap of unknown length
71847: contig of 2865 bp in length
71848: gap of unknown length
74812: contig of 3071 bp in length
74813: gap of unknown length
74913: contig of 2072 bp in length
77983: gap of unknown length
77984: contig of 1558 bp in length
80155: gap of unknown length
80156: contig of 1829 bp in length
81813: gap of unknown length
81814: contig of 2261 bp in length
81914: gap of unknown length
83743: contig of 2208 bp in length
83744: gap of unknown length
86103: contig of 1886 bp in length
86104: gap of unknown length
88411: contig of 1858 bp in length
88412: gap of unknown length
88512: contig of 1296 bp in length
90397: gap of unknown length
90398: contig of 1886 bp in length
90497: gap of unknown length
90498: contig of 1296 bp in length
92759: gap of unknown length
92760: contig of 1411 bp in length
92860: gap of unknown length
94271: contig of 2136 bp in length
94371: gap of unknown length
96507: contig of 1390 bp in length
96508: gap of unknown length
97996: gap of unknown length
97997: contig of 1858 bp in length
98096: gap of unknown length
98097: contig of 1296 bp in length
99054: gap of unknown length
99055: contig of 1296 bp in length
100054: gap of unknown length
100055: contig of 1296 bp in length
101351: gap of unknown length
101352: contig of 1886 bp in length
101451: gap of unknown length
103336: contig of 1560 bp in length
103337: gap of unknown length
104996: contig of 1560 bp in length
104997: gap of unknown length
105097: contig of 2030 bp in length
107126: gap of unknown length
107127: contig of 1442 bp in length
108668: gap of unknown length
108669: contig of 1580 bp in length
110348: gap of unknown length
110349: contig of 1870 bp in length
110448: gap of unknown length
112319: contig of 1873 bp in length
112320: gap of unknown length
114291: contig of 1873 bp in length
114292: gap of unknown length
114391: contig of 1985 bp in length
116376: gap of unknown length
116377: contig of 1859 bp in length
116476: gap of unknown length
118335: contig of 1859 bp in length
118336: gap of unknown length
118435: contig of 1512 bp in length
119947: gap of unknown length
119948: contig of 1512 bp in length
120047: gap of unknown length
121943: contig of 1896 bp in length
122048: gap of unknown length
122049: contig of 1723 bp in length
123767: gap of unknown length
123768: contig of 1168 bp in length
125034: gap of unknown length
125134: contig of 1693 bp in length
125135: gap of unknown length
126827: gap of unknown length
126828: contig of 1651 bp in length
128578: gap of unknown length
128579: contig of 1024 bp in length
129702: gap of unknown length
129703: contig of 2373 bp in length
129802: gap of unknown length
132175: contig of 1811 bp in length
132176: gap of unknown length
132276: contig of 1811 bp in length
134086: gap of unknown length
134087: contig of 1078 bp in length
134186: gap of unknown length
135264: contig of 1078 bp in length

	* 135265	135364:	gap of unknown length	
	*	135365	136997: contig of 1533 bp in length	
	*	136898	136997: gap of unknown length	
	*	136998	138452: contig of 1455 bp in length	
Query Match	35.1%;	Score 803.8;	DB 2;	Length 155623;
Best Local Similarity	68.3%;	Pred. No. 4.9e-183;	Mismatches 562;	Indels 98; Gaps
Matches 1419;	Conservative	0;		
QY	167	cgtgacgcagcagatttgaagtatcgacatacaagttgaagatcacgcctcgcaaccocg	226	
Db	1592	CGCCTCCGCGGTGGCGCGCGCATCGGTGCGCAATTCGCGCTCGCACCCGGG	1651	
QY	227	ggcctggccgggtaggggcgcgctgatcttccttccttcgccctccgcctccccct	286	
Db	1652	-----CCTCGGGGAGGGCGCGCGTCTGATTTCCTTCCTGCCCTCTGCAG---CCCC	1703	
QY	287	ggtgcgatgtccagctcaagctcgcccctcgcccttaatttatTTTTTTTcttgggcgcc	346	
Db	1704	TGTGCCATGCTCGGC-CFACCGGCCCCAGCTTGATTGA-----TCGGTCGGACGGC	1757	
QY	347	gctgcaccgggactgacctcggtggatgggaagtggagcccccggagctgctacgtggc	406	
Db	1758	GCTCGACCCTGGCGGCAGACGGCGGGATGGGAGCC-----CGGCGCTGGG	1807	
QY	407	ggcgcgctgttagagagcagccaggggagcgtcggtcgccggtgagtatccogg	466	
Db	1808	AGCGCGCAGTAGTATCAGC-----GGTGGCGCGGTGAGTAGTACCGGTGAGTACCGCGG	1859	
QY	467	aagcgcccatggggctccgtaaagaagacaccaagaacccccccgtcttcagcagga	526	
Db	1860	-----CATTGGGCTCCGCAAGAAGAACCCAGGAACCCC CGGTGTGAGCCACGA	1910	
QY	527	attcctctgcagaatcatgcggacatcgctcctcggtgggagtgcttctctgctggg	586	
Db	1911	ATTATGTTGTCAGAACACGCGGATGCTCTCTGCTGGGTGATGTTCTGTCGTGGG	1970	
QY	587	gcttgtttcagggaacagacaagatccatcgtgtttctcactcttcagcacagtgt	646	
Db	1971	ACTATTGTTTCGAGGGCAGGGCCGAGATGTCATCGTGTCTCACCCTGAGCATGGAGT	2030	
QY	647	tgctgtccctgcagcagaggaaacagccagggctcaaagtccctctatttatgtgtg	706	
Db	2031	CGTTGTCC-- -CAGCGAAGGCTACCTCGGGTCCAGGACCCTTACCATATTGGGT	2087	
QY	707	caaatgttggccacggttttcttcacatgctggtggcaatcattattccatgccacaat	766	
Db	2088	CAAAGATCTGGCCACAGTGTCTTCTACATGCTGTGGCCATCATCATACGCCACCAT	2147	
QY	767	tcagggaatgtgttgataaaaattaacaagagaatcagttcaccaagcgaacaaaa	826	
Db	2148	TCAGGAGTAGTGTAGATAAGCTCAGCGGAGACTGCAGCTACCAAAAGCAACA AAA	2207	
QY	827	caagtttaacgagctgtggttagtgtgtctactctttttctgtattttggggcac	886	
Db	2208	CAAAATTAATAGGCGGSCACTGAGTGTGTTTACATAGTGTCTGSTATCTGGGGTAT	2267	
QY	887	attcattttaactctgaaactgcctdgcagcccaactcttatatgggaagctcgtcc	946	
Db	2268	GATCATTTGGGCTCTGAGAACTGCTGTCAGACCCCACTTATTTGGGAAGTCTCAGCC	2327	
QY	947	ccaatgatgatcaatttcaaatgaagtttttacaatatcccaagttgacttacgtgtt	1006	
Db	2328	CCAACAATGATGACATTTACATGAATAATTTTCTACATCTFCACAGTTGGCTTACTTGGTT	2387	
QY	1007	tcatgtctttctgaactctactctccagaaaaaacaaaaaacagacatccctcgtcaact	1066	
Db	2388	TCATAGTTTCCCGAGCTCTACTTCCAGAAAGTCAGGAACAAGATATCCGGGTCAACT	2447	
QY	1067	tgtctacatgttgcttcaactcttccaatcaatcgtggagcttatctcttgtaactaatca	1126	
Db	2448	CATCTACATTTGCCCTCCACCTCTCCATATTTGGAGGGCCCTATCTCTTGTACTTGAACCA	2507	


```
Db 3542 ---TGACATTTAGCGAATCAAGATACCTTTAATAAATAT 3577
|||||
RESULT 8
AX329795
LOCUS AX329795
DEFINITION Sequence 304 from Patent WO0194629.
ACCESSION AX329795
VERSION AX329795.1 GI:18102773
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horligan,S., Soppet,D.R. and Weaver,Z.
Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 304 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
Location/Qualifiers
1..1267
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 343 a 278 c 275 g 371 t
ORIGIN
Query Match 32.4%; Score 742.4; DB 6; Length 1267;
Best Local Similarity 79.7%; Pred. No. 1e-168;
Matches 907; Conservative 0; Mismatches 216; Indels 15; Gaps 2;
QY 465 ggaagccaccatggggctccgttaagaagcaccacaaagaccccccgttctcagcag 524
|||||
Db 110 GCGCGCCCCACCATGCGATTGCGAAGAAAGCACCAGAGCCCCCAGTCTGAGCCAC 169
|||||
QY 525 gaattcatctgcagaatcatgcggacatgctctcgtcgtgggagttcttctcgtg 584
|||||
Db 170 GAATTCGTCCTCGAATACGCGACATCGTCTCTGTGGGATGGTCTTCTCTGCTG 229
|||||
QY 585 gggctgtgttcagggaacagcagaacatccatcgtgtttctcactcttcacacagt 644
|||||
Db 230 GGGCTCATGTTTGAGATAACGCGAAAGCTTCTATCAATTTTGTACTCTTCAGTACAAT 289
|||||
QY 645 gttgctgccctgagcagaggaacagcccggttcacaaagtcctctattattggt 704
|||||
Db 290 GTCAACCTCCCGAGCAACAGAAAGAACTACTGAATCAGTGTCCCTTTATTACTATGGC 349
|||||
QY 705 gtcaaaagatttggccacggttttctctacatgctgtggaacatcattatcagcaca 764
|||||
Db 350 ATCAAGATTGGCTACTGTCTTCTACATGCTAGTGGCGAATAATTATTCATGCCGTA 409
|||||
QY 765 attcaggaataatggttggataaaattacaagagaatgcagttcacccaaagcgaacaa 824
|||||
Db 410 ATTCAAGAGTATATGTTGGATAAAATTAACAGCGGAATGCACATTCACAAACAAACAC 469
|||||
QY 825 aacaagtttaacagagctgctcagtttagtgttcttaacttttttctgtatttggggc 884
|||||
Db 470 AGCAAGTTTAATGAATCTGTCAGCTTAGTGGCTTCTACCTTTTGTGCTGTGTTGGGC 529
|||||
QY 885 acattcatttaactctgaaaaactgctgtcagaccccaactcttatatggaagcctct 944
|||||
Db 530 ACATTTCATTCTCTCTGAAACTACATCTCAGACCCCAACTATCTATGAGGGCTTAT 589
|||||
QY 945 ccccatgacatgatgacatttcaaatgaagtttttttacatatccagttggtctactgg 1004
|||||
Db 590 CCCATAACCTGATGACATTTCAATGAAGTTTTCTACATATCAGCAGCTGGCTTACTGG 649
|||||
QY 1005 ttctatgcttttctgaactctactctccagaaaacacaaagacagatcctcgtctcaa 1064
|||||
Db 650 CTTTCATGCTTTTCTCTGAATCTACTTCCAGAAACCAAAAGAGATATTCCTCTGTGAC 709
|||||
```

```
QY 1065 cttgtctacattggtcttccaccttccacattactgagcgttactcttctgtacttgaat 1124
|||||
Db 710 CTTGTCTACATTTGGTCTTTTACCTCTTCACATTCGTGGAGCTTACCTTTTGAACCTTGAAT 769
|||||
QY 1125 catttgggaactctctcttcttggtagtgcattatttttggtaattacttcccaactgtgc 1184
|||||
Db 770 CATCTAGGACTTGTCTTCTGTGCTGCTACATATTATTTGTGTAATTTCTTTTCCACATTTCC 829
|||||
QY 1185 ggcctgttttacttttagtgaagaaagaccagaaagcattatctctgtggccattgtg 1244
|||||
Db 830 CGCCTGTTTTATTATTAGCAATGAAAGGATATCAGAAAGGATTTTCTGTGGCGAGTTCTT 889
|||||
QY 1245 ttattcttgggtagacttggtagctttaaattgttccgtactcactcactgttgggtttcactcg 1304
|||||
Db 890 TTTGTTTTGGGAAGACTTCTGACTTTAAATCTTTTTCAGTACTGACTGTGTGTTTGGCCTT 949
|||||
QY 1305 gctgagtcgagaatcggaaatcctgtagccttactggaatgtaaatgtgttggcagct 1364
|||||
Db 950 CCAAGAGCAGAAATCAGAAACTGGATTTTCAGTACTGGAACCTTCAATGTGTAGCTGTT 1009
|||||
QY 1365 aaaattgctgtctgtcgtccagttgcacgattccaaagcctacgtaacatggaacttaatt 1424
|||||
Db 1010 AGAATCGCTGTTCTGGCATCCATTTGGTTACTCAGGCATTTATGATGTGGAAGTTCAAT 1069
|||||
QY 1425 actctcgtgctcagaggtggttagaagattcttaattcagcctcctcattgataaag 1484
|||||
Db 1070 AATTTTCAGCTTCGAAGTGGAGGGAACATCTGCTTTTCAGGCACCCAGCTGTGAAGAAG 1129
|||||
QY 1485 aaacggctc-----gagatctctfaaaaaaagaagaacagaaacggagtg---gga 1529
|||||
Db 1130 AACCAACAGTACTAATAAGCAGATCTCTTAAAAAAGGAACAGAAATGCTGTGAATGA 1189
|||||
QY 1530 gtggaacttcaaatgagtagactgtccgcacaaagaggaagaaatcttctcataat 1587
|||||
Db 1190 ACATTAACTTCAATGTAGCAGACTCTCCCGGAATCTCTCCCGGAATTAANAAGAGAAATCTTCATAAT 1247
|||||
```

RESULT 9

```
BD005229 LOCUS 1267 bp DNA linear PAT 31-JAN-2002
DEFINITION Method of screening of protein.
ACCESSION BD005229
VERSION BD005229.1 GI:18633190
KEYWORDS JP 03075332-T/4.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1267)
Todo,N., Okuyama,H., Imamura,M., Ishikawa,H. and Nemoto,K.
METHOD of screening of protein
Patent: JP 03075332-T 4 16-FEB-2001;
SUMITOMO PHARMACEUTICALS CO LTD,NAOKI TODO,HAJIME OKUYAMA,OTOAKI
IMAMURA, HIRONORI ISHIKAWA, KIYOMITSU NEMOTO
COMMENT OS Homo sapiens (human)
PN JP 03075332-T/4
PD 16-FEB-2001
PF 17-AUG-2000 JP 2000005488
PR 20-AUG-1999 JP 99P 234764
PI NAOKI TODO,HAJIME OKUYAMA,MOTOAKI IMAMURA,HIRONORI ISHIKAWA,
KIYOMITSU NEMOTO
PC C1201/02,G01N33/50,C07K14/47,A61K38/17,C12N5/10,C12P21/02// PC
(C12P21/02,C12R1:91)
CC
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
1..1267
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 343 a 278 c 275 g 371 t
ORIGIN
```


Query Match	32.48;	Score	742.4;	DB	6;	Length	1267;
Best Local Similarity	79.7%;	Pred.	No. le-168;				
Matches	907;	Conservative	0;	Mismatches	216;	Indels	15;
Gaps	2;						
QY	465	ggaagcgccaccatgaggctccgtaagaagagcaccagaaccccccggttctcagccag	524				
DB	110	GGCGGCCCCACCATGGCGATTCGCAAGNAAGACCAAGAGCCCCCAGTCGCTGAGCCAC	169				
QY	525	gaattcatctgcagaatcatcgggacatogtctctcgtgctggggatgttcttctcgtg	584				
DB	170	GAATTCGTCCTGCAGAAATCAGCGGAGATCGCTCTCTGCTGTGGGATGGTCTTCTGCTG	229				
QY	585	ggccttggttcgaggaaacagcagaagcatccatcgtgtttctcactcttcacacag	644				
DB	230	GGGCTCATGTTTGAGATTAACGGCAAAAGCTTCTATCATTTTTTGTGTACTCTTCAGTACAA	289				
QY	645	gttgctgcctgcagcagagggaacgacggcggcctcaagtcacctctattattatggt	704				
DB	290	GTCACTCCAGCAACAGAAAGAAAGCTACTGAATCAGTGTCCCTTTATTACTATGGC	349				
QY	705	gtcaagatctggccacgggtttctctacatgctggtggcaatcattattatcagcaaca	764				
DB	350	ATCAAAAGATTGGCTACTGTTTCTTCTACATGTAGTGGCGATAATTTATTCATGCCGTA	409				
QY	765	attcaggaatatgtgttggaataaatacaacagagaatcgagttcaccacagcgaacaa	824				
DB	410	ATTCAAGAGTATATGTGGATAAATTAACAGGGCGAATGCATCTCCNAACAAACAC	469				
QY	825	acaagtttaacgagctggtgcagtttagtgtgtctacttttttttctgtattggggc	884				
DB	470	AGCAAGTTTAATGAATCTGCTGAGCTTAGTGGCTTACCTTTTTTGTGTTGGGGC	529				
QY	885	acattcattttaactctgaaactgcgtgcagaccacactctatatgaagcctgt	944				
DB	530	ACATTCATTCTACTCTGAAACTACATCTCAGCCCACTATCTTATGGAGGGCTTAT	589				
QY	945	cccataagcatgatgcatttccaaatgaagtgttttctacatcattccagttgggttactgg	1004				
DB	590	CCCATAACCTGATGACATTTCAATGAAGTTTCTACATATCACAGCTGGCTTACTGG	649				
QY	1005	ttcatcgttttcttgactctacttccagaaacccaaaaaacagacatccctcgtaac	1064				
DB	650	CTTCATGCTTTCTCGTAACCTACTTCCAGAAACACCAAAAGAAAGATATTCCTCGTCG	709				
QY	1065	cttgctacatgggtcttcacctctccacattactgagcgttatctctgttacttgtaat	1124				
DB	710	CTTCTCTACATTGCTCTTTACCTCTCCACATTGCTGGAGCTTACCTTTTGAACCTGAAT	769				
QY	1125	catttgggaactctcttttggtaactgcattatttttgaattacttccacatgtgc	1184				
DB	770	CATCTAGGACTGTGTCTCTGGTGCTACATTTATTTGTGAATTTCTTTTCCACATTTCC	829				
QY	1185	ggcctgttttactttagtgatgaaggtacagaaagcatatctctgtggggcattgtg	1244				
DB	830	CGCCTGTTTTATTTAGCAATGAAAGATACAGAAAGGATTTTCTCTGTGGGCGAGTCTT	889				
QY	1245	tttatcttgggtagactgtgactttaattgtttccgtactcactggttgggtttcaactg	1304				
DB	890	TTTGTGTTGGGAACACTTCTGACTTTAATCTTTTTCAGTACTGACTGTGTGGTTTTGGCCTT	949				
QY	1305	gctggatgcagaatcgggaatcctgatgccccttactggaaaatgtaaatgtgttggcagct	1364				
DB	950	GCAAGACGAGAAAATCAGAAACTGGATTTTCAGTACTGGAAACTTCAATGTGTTAGCTGT	1009				
QY	1365	aaaattgctgtctcgtccagctgcacgatccagctcaagctacgttaaacatgaaacttaatt	1424				
DB	1010	AGAACTCGCTTCTCGGCATCCATTTTGGCTTACTCAGGCATTTTATGATGTGGAAGTTCAAT	1069				
QY	1425	actctcgtgttcagaggtgggtagagattcttaatatccaggcctcatgtatgaaaaag	1484				
DB	1070	AATTTTCAGCTTCCGAAGTGGAGGAACTACTGCTTTTCTTTCAGGCACCACTGTGGAAG	1129				

Qy	1485	aaacggtc-----gagatcttctataaaagagacagaaaaacggagtg---gga	1529
Db	1130	AAAAACCAACAGTAACATAAAGCGCAGATCTTCTTAAAAAGGACAGAAAAATGGTGTGAATGGA	1189
Qy	1530	gtggaaccttcaaatagatgactgctccgcaagaggaagaaatcttcataat	1587
Db	1190	ACATTAACTTCAATGTAGCAGACTCTCCCGGAAATAAAAAAGAGAAATCTTCATAAT	1247
RESULT	10		
HSTRAMP			
LOCUS		1267 bp	linear
DEFINITION	H.sapiens mRNA for TRAMP protein.		
ACCESSION	X63679		
VERSION	X63679.1 GI:37264		
KEYWORDS	tram gene; tram protein.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1267)		
JOURNAL	Gorlich,D., Hartmann,E., Prehn,S. and Rapoport,T.A.		
MEDLINE	A protein of the endoplasmic reticulum involved early in		
REFERENCE	polypeptide translocation		
AUTHORS	Nature 357 (6373), 47-52 (1992)		
JOURNAL	92244357		
MEDLINE	2 (bases 1 to 1267)		
REFERENCE	Hartmann,E.		
AUTHORS	Direct Submission		
TITLE	Submitted (28-JAN-1992) E. Hartmann, Max-Delbrueck-Centr. f.		
JOURNAL	Molekulare Med., Robert-Rössle-Strasse 10, O-1115 Berlin Buch, FRG		
FEATURES	Location/Qualifiers		
source	1..1267		
gene	/organism="Homo sapiens"		
CDS	/db_xref="taxon:9606"		
	/cell_line="HeLa"		
	122..1246		
	/gene="TRAM"		
	122..1246		
	/gene="TRAM"		
	/codon_start=1		
	/product="TRAM protein"		
	/protein_id="CAA45218.1"		
	/db_xref="GI:37265"		
	/db_xref="SPREMBL:Q15629"		
	/translation="MAIRKSTSPVLSHEFVLQNHADIVSCVAMVFLGLMFETA		
	KASTIFLYQNTVLPKATQAEVSLVYVGKIDATLAFYMLVATIIHAYIQEYML		
	DKNRRMHFSKTHKSPNESGSLSAFLFACVNGTFILISENVISDPTILWRAYPHNL		
	MTQMKEFYISQLAYWLHAPPELYFQKKEDIPRLQVLYLFIHFTAGAYLLNLHL		
	GLVLLVYEFELFHSRLYFESNEKIQKGLSLWAVFLVIGRLLLTLLISLVLTGFGG		
	ARAENQLKIDSTGNFNVAIRIYALVASICIQAPMMKKFINFOLRRRREHSAFOAPAA		
	KKPKVTIKGRSSKKTENGNGVLTLSNVADSPRNKKEKS"		
BASE COUNT	343 a 278 c 275 g 371 t		
ORIGIN			
Query Match	32.4%; Score 742.4; DB 9; Length 1267;		
Best Local Similarity	79.7%; Pred. No. 1e-168;		
Matches	907; Conservative 0; Mismatches 216; Indels 15; Gaps 2;		
Qy	465	ggaagcgccaccatggggctcgtaagaagagaccacgaaccccccgcttcagccag	524
Db	110	GGCGGCCCCACATGGCGGATTTCGGAAGAAAAGACACCAAGAGCCCCCAGTCTGAGCCAC	169
Qy	525	gaattcatctcgagaatcatcgcgacatcgctctctcgctggggatgtcttctctgctg	584
Db	170	GAATTCGTCCTCGAGAAATCACGCGACATCGTCTCCTGTGTGGCATGGTCTTCTCTG	229
Qy	585	gggcttggttcgaggggaacgagagaatccatcggtttctctactcttcagcaggt	644
Db	230	GGGCTCATGTTTGAGATAACGCGAAAAGCTTCTCATTTTGTGTACTCTTCAGTACAAT	289

Query Match 32.4%; Score 742.4; DB 9; Length 1267;
Best Local Similarity 79.7%; Pred. No. 1e-168;
Matches 907: Conservative 0; Mismatches 216: Indels 15; Gaps 2;

[illegible]

Qy	645	gttgctgtccctgcagcagaggaacagccagcggtccaaagtcocctctattattatggt	704
Db	290	GTACCCCTCCAGCAACAGAGAACAGCTACTGAATCAGTGTCCCTTATTACTATGGC	349
Qy	705	gtcaaaagatttggccagcggtttttcttctacatgctggtggcaatcattattcatgcccaca	764
Db	350	ATCAAAAGATTGGCTACTGTTTCTTCTACATGCTAGTGGCGATAAATTATTTCATGCCGTA	409
Qy	765	attcaggaatatgtgttgataaaattacaagaagatgcagttccacaaagcgaacaa	824
Db	410	ATTCAAGAGTATATGTTGGATAAAATTAACAGGCGAAGCACTTCTCCAAAACAAACAC	469
Qy	825	aacaagtttaacagagtcgtgcagtttagtgtcttactttttttcttatttatttggggc	884
Db	470	AGCAAGTTTAATGAATCTGGTCACGTTAGTGGCTTCTACCTTTTGGCTGTGTTGGGGC	529
Qy	885	acattcatttttaactctgtaaaactgcctgtcagaccacaaactcttatatggaaggtcgt	944
Db	530	ACATTCAATCTCATCTCTGAAACTACATCTCAGACCCAACTATCTTATGGAGGCTTAT	589
Qy	945	ccccatgacatgacatttcaaatgaagtttttctacatatcccaagttggcttactgg	1004
Db	590	CCCATACCTTGATGACATTTCAANTGAAGTTTCTTACATATCACAGCTGGCTTACTGG	649
Qy	1005	tttcatgcttttccctgaactctacttccagaaacccaaaaaacagacatccctcgtcaa	1064
Db	650	CTTCATGCTTTCTCTGAACCTACTTCCAGAAAACCAAAAAGAAAGATATCTCTCGTCAG	709
Qy	1065	cttgctacatgtgcttcaoctcttccacattactgtgagcttatctcttctgttaoctgaa	1124
Db	710	CTTGCTACATTTGCTTTTACCTCTCCACATTTGCTGGAGCTTACCTTTTGAACCTTGAAT	769
Qy	1125	catttgggaactctcttcttggactcattatttttgaattactttccacatgtgc	1184
Db	770	CATCTAGACCTTGCTCTCTGGTGCTACATATTTTGTGAAATTTCTTTCCACATTTCC	829
Qy	1185	ggcctgttttacttttagtgaagaatgaccagaaagccatctctgtggggcattgtg	1244
Db	830	CGCCTGTTTTATTATTTAGCAATGAAAGTATCAGAAAGGATTTTCTCTGTGGCAGTTCTT	889
Qy	1245	tttatcttggtagacttgaactttaattgtttcccgtaactcactgctgtggttttcaactg	1304
Db	890	TTTGTGTTGGGAACACTCTGACTTTAATCTTTTCAGTACTGACTGTGTGTTTGGCCTT	949
Qy	1305	gctgagtcgacagaatccctgtagcccttactggaatgtaaatgtgttggcagct	1364
Db	950	GCAAGACAGAAATCAGAACTGGATTTTCAGTACTGGAAACTTCAATGTGTAGCTGTT	1009
Qy	1365	aaaattgctgtctgtcgtccagttgacagatccaaagccctacgtacatggaacttaatt	1424
Db	1010	AGAAATCGTGTCTTGGCATCCATTTGGGTTACTCAGGCATTTATGATGTGGAAGTTT	1069
Qy	1425	actctcgttctcagaggtgggtagagattctaattcagccctcatgtatgaaaaag	1484
Db	1070	AATTTCAGCTTCTGAAGGTGGAGGACATCTGCTTTTCAGGCACCAGCTGTGGAAGAAG	1129
Qy	1485	aaacggctc-----gagatctcttaaaaaaagacagaaacggagtg---gga	1529
Db	1130	AAACCAACAGTAACTAAGGCAGATCTTCTTAAAAAAGAACAGAAAATGTGTGAATGA	1189
Qy	1530	gtggaacttcaaatagtagactgtccgcaagaggaaagagaataatttccataat	1587
Db	1190	ACATTAACTTCAAAATGTAGCAGACTCTCCCGGGAATAAAAAAGAAATCTTCATAAT	1247
RESULT 11			
AY029764			
LOCUS	2720 bp	mRNA	linear
DEFINITION	Mus musculus TRAM1 mRNA, complete cds.		
ACCESSION	AY029764		
VERSION	AY029764.1		
KEYWORDS	house mouse.		
SOURCE			

ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
AUTHORS	Hartmann, E.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-APR-2001) Biologie, Universitaet Luebeck, Ratzeburger Allee 160, Luebeck 23538, Germany		
FEATURES	Location/Qualifiers		
source	1. .2720		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="R2PD clone IMAGp98K219524g2"		
	36. .1160		
CDS	/note="Tram1"		
	/codon_start=1		
	/product="TRAM1"		
	/protein_id="AAK38167.1"		
	/db_xref="GI:13774095"		
	/translation="MAIRKSKNPPLLSHEFLONHADIVSCLAMFLFLGLMFEVTA		
	KGAIIPVALOYNTVPATEQATSESLYHYGIKDLATVLYMVAIIIIHAIIOEVVL		
	KINRMHESKTKHSPNESGQLSAPYLFACVWGFIILSENISDPTILRAYPNHL		
	MTQNRFFYSIQLAYWLHAFPEIDYFQTKKEDIIPOLVYIGLYFIHAGIAYLLNHL		
	GVLLVLFHYFVLEFHSIRLFYSEKQKFSLWALFVLGLRLTLILSLVTVGFL		
	ARAENOKLDFSTGNFNLAVRIAVLASICITQAFMMKTFINQLRRWREHSFAQAPPV		
	KRKPAVTKGRSSRKGTENGVTGTVTSNGADSPNRKESK"		
BASE COUNT	722 a	569 c	899 t
ORIGIN			
Query Match	32.0%;	Score 733;	DB 10; Length 2720;
Best Local Similarity	74.0%;	Pred. No. 2.3e-166;	
Matches	1005;	Conservative	0; Mismatches 330; Indels 24; Gaps 5;
Qy	465	ggaagcgccaccatgggggtccctgaagaagagcaccacgaaccccccgctctccagccag	524
Db	24	GGCGGCTCCACCATGGCGATTCGCAAGAGCAAGCAACGCCGCGCTGCTGAGCCAC	83
Qy	525	gaattcatcctcagaaatcgcgacatgctctcgtgggagtgcttcttctcgtcg	584
Db	84	GAATTTCTGCTGAGAAATCAGCGGACATGCTCTCTGCTGGCGATGCTTTCTCTGCTG	143
Qy	585	ggcctgtgtctcagggaaacagcagaagcatccatcgtgtttctcactcttccagcag	644
Db	144	GGCTTATGTTTCGAGGTAAACAGCAAAAGGAGCCATCATCTTTGCTCTTCAGTATAAT	203
Qy	645	gtcgtcgtccctcagcagaggaacacagccacggtccaaagtcctctattattatggt	704
Db	204	GTTACCGCGCCCTCAACAGAAAGCAAGCTACTGAAATCAGCATCCCTATATCATATG	263
Qy	705	gtcaaaagatttggccaggttttctctacatgctggtggcaatcattatccatgccaca	764
Db	264	ATCAAGACTTGGCTACAGTTTGTCTACATGCTAGTGGCGATAAATTATTTCATGCCATA	323
Qy	765	attcaggaatatgtgttggataaaattacaagaagaatgcagttccaccaagcgaacaa	824
Db	324	ATTGAGGAGTATGTTGGGATAAAATTAACCGCGGATGCATCTTCCAAAACGAAGCAC	383
Qy	825	aacaagtttaacagagtcgtgcagtttagtgttcttacttttttttcttatttggggc	884
Db	384	AGCAAGTTTAAACAGAGCTGGTGCAGTCAGTCATCTTACCTTTTGGCTGTGTGGGGC	443
Qy	885	acattcatttttaactctctgaaactgcctgcagaccacaaactcttatatggaaggtcgt	944
Db	444	ACGTTTCTATCTCCGAAAACACTATATCTCAGACCAACTATCTCTGGAGGCGCTAT	503
Qy	945	ccccatgacatgacatttcaaatgaagtttttctacatatccccagttgggttactg	1004
Db	504	CCCATAAACCTGATGACATTTTCAGACGAAGTTTTTCTTACATCTCCAGCTGGGTACTGG	563
Qy	1005	tttcatgcttttccctgaactcttacttccagaaaaaaccaaacagacatccctcgtcaa	1064
Db	564	CTCCATGCATTTCCCTGAACCTCTACTTCCAGAAAACCAAAAAGAAATATATCCCTCGTCAG	623

424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

REMARK NIH-MGC Project #

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS	1 (bases 1 to 2819)
TITLE	Strausberg, R.
JOURNAL	Direct Submission
REMARK	Submitted (15-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA NIH-MGC Project URL: http://mgc.nci.nih.gov

Qy	825	aaacagtttaacagagctggtgcagtttagtgtgtctacttttttctgttattgttgggc	884	DEFINITION	Homo sapiens chromosome 4 clone RP11-724E21 map 4, LOW-PASS
Db	481	ACGAAGTTTACGAGCTCGTGCAGCTAGTCATCTACCTTTTTCCTGCTGTGGGGC	540	ACCESSION	AC027246
Qy	885	acattcatttaactctgaaacagctgcgtgcagaccacactcttatatgaaggtcgt	944	VERSION	AC027246.1 GI:7331616
Db	541	ACGTTCAATTCATCTCCGAAACATATATCTCAGACCCCAACTATCTGTGAGAGCCAT	600	KEYWORDS	HTG; HTGS_PHASE0.
Qy	945	ccceatagcatgatacattcaaatgaagtttttctacatatccacagttggcttactgg	1004	SOURCE	human.
Db	601	CCCCATAACCTGATGACATTTTCAGACGAAGTTTTCATATCTCCACCTGGCGTACTGG	660	ORGANISM	Homo sapiens
Qy	1005	tttactgttttctgaaactactctcagaaaaacacacacacacacacacacacacac	1064	REFERENCE	1 (bases 1 to 73339)
Db	661	CTCCATGATATCCCTGAACTCTACTTCCAGAAACCAACCAAGAGATATCCCTGCTCAG	720	AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galaan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lacombe, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McArthur, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melclrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T. M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Qy	1305	gctgagatcgagaatcgatccctactgacatgaaatgtaaatgtttggcagct	1364	TITLE	Direct Submission
Db	961	GCAGAGCGGAGAAFCAGAGCTGGACCTTGTACTGGAAATTCACAGCTGTGGCCGTT	1020	JOURNAL	Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
Qy	1365	aaattgtgttctgtcgtccagttgacgactcgaagctacgtaacatggaacttaatt	1424	COMMENT	All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html
Db	1021	AGGATCGCTGTTCGCGCTCCATTTGCATCACACAGCCCTTCATGATGTGGAAGTTTCAT	1080	Center: Whitehead Institute/ MIT Center for Genome Research	
Qy	1425	actctctggtctcagagtggtgagaagattcctaataatcaggctctatgatgaaaaag	1484	Center code: WIBR	
Db	1081	AACTTCCAGCTCGGAGGTGGAGGAACATTTGCCTTTCCAGGCCCCACCTGTGAAAAGG	1140	Web site: http://www-seq.wi.mit.edu	
Qy	1485	aaacggtc-----gagatcttctaaaaaagacagaaaaacggagtg---gga	1529	Contact: sequence.submissions@genome.wi.mit.edu	
Db	1141	AAACCGCGCTGACCAAGGAGGTCTTCCAGAAAGGAACAGAGACGGTGTGAACGGC	1200	Project Information	
Qy	1530	gtggaaactcaaatagagtagactgtcgcgcaagagagaaagaaatcttccataatct	1589	Center project name: L8852	
Db	1201	ACAGTCACATCAACAGGAGCAGACTCCCTCGGAACAGGAAAGAGAGCTTCGTAAATGG	1260	Center clone name: 724_E_21	
Qy	1590	ttgcaagcgcattgatttaattctgcaagaatctgctcttggaggtttctttctgtcac	1649	* NOTE: This record contains 82 individual	
Db	1261	ATCGGAAGCTGACTGATTAATTATTTGCCCAAGAAATCTGCTTTCTGCTCTATCTCAGCAC	1320	* sequencing reads that have not been assembled into	
Qy	1650	tagagatttttctgtttttgaaa-atagttcgtctctctggtttttgtttattgaactgt	1708	* contigs. Runs of N are used to separate the reads	
Db	1321	TAGCAATTTTCTTTCTFAAAATACAGTTATGCTCTTTGATTCTCTGCTGTTGCCGT	1380	* and the order in which they appear is completely	
Qy	1709	ttcatgtatttttaaaacacattttgaggggagggattattatgaatgggaaaaaaga	1768	* arbitrary. Low-pass sequence sampling is useful for	
Db	1381	TCC---TGCATCGTAAAGGCGATTGAGGGAGGATATACATCTGAGTG-----AGACA	1432	* identifying clones that may be gene-rich and allows	
Qy	1769	ttttggttgagactaaataactaatcgtcgaataatgt	1807	* overlap relationships among clones to be deduced.	
Db	1433	TTGTAGCTTAGACTACAGCTACCTGCTTCCATTCAGAAATAGTTT	1471	* However, it should not be assumed that this clone	
RESULT 13				* will be sequenced to completion. In the event that	
AC027246/c				* the record is updated, the accession number will	
LOCUS				* be preserved.	
				* 1 775: contig of 775 bp in length	
				* 776 875: gap of 100 bp	
				* 876 1673: contig of 798 bp in length	
				* 1674 1773: gap of 100 bp	
				* 1774 2552: contig of 779 bp in length	
				* 2553 2652: gap of 100 bp	
				* 2653 3464: contig of 812 bp in length	
				* 3465 3564: gap of 100 bp	
				* 3565 4277: contig of 713 bp in length	
				* 4278 4377: gap of 100 bp	
				* 4378 5176: contig of 799 bp in length	
				* 5177 5276: gap of 100 bp	

```

* 5277 6071: contig of 795 bp in length
* 6072 6171: gap of 100 bp
* 6172 6952: contig of 781 bp in length
* 6953 7052: gap of 100 bp
* 7053 7844: contig of 792 bp in length
* 7845 7944: gap of 100 bp
* 7945 8722: contig of 778 bp in length
* 8723 8822: gap of 100 bp
* 8823 9633: contig of 811 bp in length
* 9634 9733: gap of 100 bp
* 9734 10513: contig of 780 bp in length
* 10514 10613: gap of 100 bp
* 10614 11422: contig of 809 bp in length
* 11423 11522: gap of 100 bp
* 11523 12315: contig of 793 bp in length
* 12316 12415: gap of 100 bp
* 12416 13201: contig of 786 bp in length
* 13202 13301: gap of 100 bp
* 13302 14093: contig of 792 bp in length
* 14094 14193: gap of 100 bp
* 14194 14978: contig of 785 bp in length
* 14979 15078: gap of 100 bp
* 15079 15877: contig of 799 bp in length
* 15878 15977: gap of 100 bp
* 15978 16781: contig of 804 bp in length
* 16782 16881: gap of 100 bp
* 16882 17669: contig of 788 bp in length
* 17670 17769: gap of 100 bp
* 17770 18566: contig of 797 bp in length
* 18567 18666: gap of 100 bp
* 18667 19473: contig of 807 bp in length
* 19474 19573: gap of 100 bp
* 19574 20373: contig of 800 bp in length
* 20374 20473: gap of 100 bp
* 20474 21273: contig of 800 bp in length
* 21274 21373: gap of 100 bp
* 21374 22175: contig of 802 bp in length
* 22176 22275: gap of 100 bp
* 22276 23074: contig of 799 bp in length
* 23075 23174: gap of 100 bp
* 23175 23965: contig of 791 bp in length
* 23966 24065: gap of 100 bp
* 24066 24866: contig of 801 bp in length
* 24867 24966: gap of 100 bp
* 24967 25772: contig of 806 bp in length
* 25773 25872: gap of 100 bp
* 25873 26685: contig of 813 bp in length
* 26686 26785: gap of 100 bp
* 26786 27577: contig of 792 bp in length
* 27578 27677: gap of 100 bp
* 27678 28470: contig of 793 bp in length
* 28471 28570: gap of 100 bp
* 28571 29373: contig of 803 bp in length
* 29374 29473: gap of 100 bp
* 29474 30223: contig of 750 bp in length
* 30224 30323: gap of 100 bp
* 30324 31114: contig of 791 bp in length
* 31115 31214: gap of 100 bp
* 31215 32038: contig of 824 bp in length
* 32039 32138: gap of 100 bp
* 32139 32951: contig of 813 bp in length
* 32952 33051: gap of 100 bp
* 33052 33845: contig of 794 bp in length
* 33846 33945: gap of 100 bp
* 33946 34753: contig of 808 bp in length
* 34754 34853: gap of 100 bp
* 34854 35662: contig of 809 bp in length
* 35663 35762: gap of 100 bp
* 35763 36571: contig of 809 bp in length
* 36572 36671: gap of 100 bp
* 36672 37457: contig of 786 bp in length
* 37458 37557: gap of 100 bp
* 37558 38344: contig of 787 bp in length

```

```

* 38345 38444: gap of 100 bp
* 38445 39246: contig of 802 bp in length
* 39247 39346: gap of 100 bp
* 39347 40118: contig of 772 bp in length
* 40119 40218: gap of 100 bp
* 40219 41012: contig of 794 bp in length
* 41013 41112: gap of 100 bp
* 41113 41915: contig of 803 bp in length
* 41916 42015: gap of 100 bp
* 42016 42808: contig of 793 bp in length
* 42809 42908: gap of 100 bp
* 42909 43708: contig of 800 bp in length
* 43709 43808: gap of 100 bp
* 43809 44604: contig of 796 bp in length
* 44605 44704: gap of 100 bp
* 44705 45507: contig of 803 bp in length
* 45508 45607: gap of 100 bp
* 45608 46403: contig of 796 bp in length
* 46404 46503: gap of 100 bp
* 46504 47319: contig of 816 bp in length
* 47320 47419: gap of 100 bp
* 47420 48211: contig of 792 bp in length
* 48212 48311: gap of 100 bp
* 48312 49111: contig of 800 bp in length
* 49112 49211: gap of 100 bp
* 49212 49925: contig of 714 bp in length
* 49926 50025: gap of 100 bp
* 50026 50816: contig of 791 bp in length
* 50817 50916: gap of 100 bp
* 50917 51719: contig of 803 bp in length
* 51720 51819: gap of 100 bp
* 51820 52609: contig of 790 bp in length
* 52610 52709: gap of 100 bp
* 52710 53505: contig of 796 bp in length
* 53506 53605: gap of 100 bp
* 53606 54387: contig of 782 bp in length
* 54388 54487: gap of 100 bp
* 54488 55295: contig of 808 bp in length
* 55296 55395: gap of 100 bp
* 55396 56184: contig of 789 bp in length
* 56185 56284: gap of 100 bp
* 56285 57082: contig of 798 bp in length
* 57083 57182: gap of 100 bp
* 57183 57992: contig of 810 bp in length
* 57993 58092: gap of 100 bp
* 58093 58890: contig of 798 bp in length
* 58891 58990: gap of 100 bp
* 58991 59775: contig of 785 bp in length
* 59776 59875: gap of 100 bp
* 59876 60669: contig of 794 bp in length
* 60670 60769: gap of 100 bp
* 60770 61575: contig of 806 bp in length
* 61576 61675: gap of 100 bp

```

```

Query Match      28.3%; Score 648.2; DB 2; Length 73339;
Best Local Similarity 99.5%; Pred. No. 1.8e-145;
Matches 650; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1636 gttctttctgcactagagattttctgttttttgaagaaagtcgtcgtcgtcgttttt 1695
DB 71476 GGTTCCTCTGCACCTAGAGATTTTCTGTTTTTGAAGAAATAGTTCGTGCTCTTCGGTTTT 71417
QY 1696 gttattgaactgttcattgtatttttaagacatttgaggagagattattatga 1755
DB 71416 GTATTGAACTGTTTCATCTATTTTTTTAAAGACATTTTGAAGGAGGAGGATTTATGA 71357
QY 1756 tgggaaaaaagattttggttgagactaaattactcatcgtcaaaataatgtcaaatag 1815
DB 71356 TGGGAAAAAAGATTTTGGTTGAGACTAAATTTACTCATCTGTCACAAATATGTCACAAATAG 71297
QY 1816 ttttggggatcaccaactatattttgttggatttttaacctttcaacatttctctaata 1875
DB 71296 TTTTGGGGATCACCACTATATTTTGTGTTTTTAACTTTTAACTTTTCAACATTTTCTCTAATGA 71237

```

```
QY 1876 ttgcagagataactgcacaattttgcatatcaatgatgactggtttactccaccagt 1935
Db 71236 TTTGCAGAGATAACTGCACAAATTTTCATATCAATGATGCTGTTACTCCACCAGT 71177
QY 1936 gtttcataataactaacagatggtctctcctagcaagattatggtttaatgcttgttt 1995
Db 71176 GTTTCATAAATACTAACAGATGCTCTCTCTAGCAAGATTATGTTTAAATGCTGCTTT 71117
QY 1996 ggggtataataaaagtcacgaaaaagtggaagtcacaaatcagatctctgttaattgtagaa 2055
Db 71116 GGGGTAAATAAAGTAGCAAAAAGGTGGGAAGTCAAATCAGTATCTGTAAATTTGTAGAA 71057
QY 2056 tttatttttaagaacttaacaactcagaaaagattgctagactcaccacaaataataaagt 2115
Db 71056 TTTATTTTAAAGAACTTACAACCTCAGAAAAGATTGCTAGACTCACCAAAATTAATAATG 70997
QY 2116 tctttattttacaggttagtattattagtgcttcattcccccatttaaaaaaacacagtac 2175
Db 70996 TTCTTTATTTTACAGGTAGTGATATTAGTGTCTATCCCCATTTAAAAAACACAGTAC 70937
QY 2176 taatgggtaacacatatggaggttgcgcacatatattgcatcaaaataatcaatt 2235
Db 70936 TAATGGGTAAACATATGGAGGTTTGTGCTGCATATATATGATCAAAATATCAATTAAT 70877
QY 2236 aatataaaaataaaaattcctcctgctccacttccacttgaatgggaattc 2288
Db 70876 AATATAAAAATAATTAATAATCTTTCTGTCTCATTCCTCCCTGTTAAATGGGAATTC 70824

RESULT 14
BD005227
LOCUS BD005227 1092 bp DNA linear PAT 31-JAN-2002
DEFINITION Method of screening of protein.
ACCESSION BD005227
VERSION BD005227.1 GI:18633188
KEYWORDS JP 03075332-T/2.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1092)
Todo,N., Okuyama,H., Imamura,M., Ishikawa,H. and Nemoto,K.
Method of screening of protein
Patent: JP 03075332-T 2 16-FEB-2001;
SUMITOMO PHARMACEUTICALS CO LTD,NAOKI TODO,HAJIME OKUYAMA, OTOAKI
IMAMURA, HIRONORI ISHIKAWA, KIYOMITSU NEMOTO
OS Rattus sp. (rat)
PN JP 03075332-T/2
PD 16-FEB-2001
PF 17-AUG-2000 JP 2000005488
PI 20-AUG-1999 JP 99P 234764
PI NAOKI TODO,HAJIME OKUYAMA,MOTOAKI IMAMURA,HIRONORI ISHIKAWA,
PI KIYOMITSU NEMOTO
PC C12Q1/02,G01N33/50,C07K14/47,A61K38/17,C12N5/10,C12P21/02// PC
(C12P21/02,C12R1:91)
CC
FH Key Location/Qualifiers
FT CDS (i). (1089).
FEATURES
source Location/Qualifiers
1..1092
/organism="Rattus sp."
/db_xref="taxon:10118"
BASE COUNT 260 a 268 c 288 g 276 t
ORIGIN

Query Match 28.3%; Score 647.8; DB 6; Length 1092;
Best Local Similarity 76.1%; Pred. No. 7.5e-146;
Matches 844; Conservative 0; Mismatches 247; Indels 19; Gaps 3;

QY 477 atggggtcgtgaagagaccacaaagacccccccgttctcagcagggaattcctg 536
```


.....

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 14:57:08 ; Search time 495.6 Seconds
(without alignments)
7926.352 Million cell updates/sec

Title: US-09-807-470-3

Perfect score: 2288

Sequence: 1 tataggcgcacgtgtgta.....ccactgttaaatggaattc 2288

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

```

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	DB ID	Description
1	2288	100.0	2288 21 AAA38013	Human WAR-1 nucleo
2	1808.8	79.1	1835 23 AAS76421	DNA encoding novel
3	1734.4	75.8	1736 22 AAK94181	Human full-length
4	1348.2	58.9	1517 24 AAS18576	CDNA encoding huma
5	1288	56.3	1288 22 AAF74782	Human WAR-1 encodi
6	1110	48.5	1110 22 AAF74780	Human WAR-1 encodi
7	937.6	41.0	2833 21 AAC77810	Human cancer assoc
8	923.4	40.4	2756 24 AAS62269	CDNA sequence #56
9	802.2	35.1	2311 21 AAA38012	Rat WAR-1 nucleoti

10	761.6	33.3	802	22	AAK91674	Human cDNA 5'-end
11	761.6	33.3	802	22	AAK93837	Human cDNA clone r
12	679.2	29.7	1120	20	AAZ33531	Human prostate can
13	647.8	28.3	1092	22	AAF74781	Rat WAR-1 encoding
14	504.4	22.0	545	22	AAK92492	Human cDNA 3'-end
15	486.8	21.3	642	22	AAS47389	Human breast cance
16	486.8	21.3	642	22	AAF17959	Human breast cance
17	298.8	13.1	1380	22	AAI01678	Human reproductive
18	273	11.9	472	21	AAC00195	Human secreted pro
19	161	7.0	1526	23	ABL18595	Drosophila melanog
20	161	7.0	1549	23	ABL18061	Drosophila melanog
21	161	7.0	2153	23	ABL21096	Drosophila melanog
22	161	7.0	3346	23	ABL21098	Drosophila melanog
23	161	7.0	3618	23	ABL18060	Drosophila melanog
24	161	7.0	4146	23	ABL18594	Drosophila melanog
25	161	7.0	4775	23	ABL21091	Drosophila melanog
26	161	7.0	5055	23	ABL21094	Drosophila melanog
27	161	7.0	11287	23	ABL21090	Drosophila melanog
28	119	5.2	447	22	AAI28950	Colon tumour relat
29	93.8	4.1	271	22	AAI23061	Human breast cance
30	92.2	4.0	205	16	AAI19252	Human gene signatu
31	91.4	4.0	272	22	AAI14199	Human breast cance
32	88.4	3.9	494	24	ABI99445	Mouse ischaemic co
33	85.6	3.8	349	21	AAC06201	Human secreted pro
34	65.6	2.9	602	22	AAK60814	Human cancer agent
35	58.4	2.6	3641	23	AAS88713	DNA encoding novel
36	58.4	2.6	6475	23	AAK91168	DNA encoding novel
37	58.4	2.6	10771	23	AAS85906	DNA encoding novel
38	54.8	2.4	6478	22	AAK45416	Chemically pretrea
39	54.8	2.4	6626	22	AAS46809	Tumour suppressor
40	52.6	2.3	7498	24	ABL32257	Human immune syste
41	52.4	2.3	16766	24	ABL34157	Human immune syste
42	52.2	2.3	6106	22	AAS46430	Tumour suppressor
43	52.2	2.3	6106	22	ABL33473	Human immune syste
44	52	2.3	15416	24	ABL34231	Human immune syste
45	52	2.3	15416	24	AAS61453	Human gene regulat

ALIGNMENTS

RESULT 1
AAA38013
ID AAA38013 standard; DNA; 2288 BP.
XX
AC AAA38013;
XX
DT 22-AUG-2000 (first entry)
XX
Human WAR-1 nucleotide sequence.
XX
DE Endoplasmic reticulum; WAR-1; cancer cell proliferation inhibitor;
XX
KW diagnosis; cancer; sarcoma; human; ds.
XX
OS Homo sapiens.
XX
PN WO200022123-A1.
XX
PD 20-APR-2000.
XX
PF 13-OCT-1999; 99WO-JP05631.
XX
PR 13-OCT-1998; 98JP-0290711.
XX
PA (SUMO) SUMITOMO PHARM CO LTD.
XX
PI Tohdoh N, Yoshima T, Komiya K, Tojo S, Nemoto K, Ishikawa H;
XX
PI Okuyama H;
XX
XX WPI; 2000-317980/27.
DR P-PSDB; AAY98147.
XX
PT Endoplasmic reticulum protein WAR-1 which inhibits cancer cell

Db 1861 acatttcttaattgattgcagagataactgcacaaattttgcatatcaatgatctggtt 1920
QY 1921 cttactccaccagtggtttcataactaactaagaatggtctctcctcagcaagattatgtg 1980
Db 1921 cttactccaccagtggtttcataactaactaagaatggtctctcctcagcaagattatgtg 1980
QY 1981 tttaatgcttctgttgggttaaaataaaagtacgaaaaaggtggaagtcaaatcagttat 2040
Db 1981 tttaatgcttctgttgggttaaaataaaagtacgaaaaaggtggaagtcaaatcagttat 2040
QY 2041 ctgtaattgtagaatttttttaagaacttacaaactcagaaaaagattgctagactca 2100
Db 2041 ctgtaattgtagaatttttttaagaacttacaaactcagaaaaagattgctagactca 2100
QY 2101 ccaaaaataaaatgctttttttttacaggtagtgattattagtgcttcacatcccat 2160
Db 2101 ccaaaaataaaatgctttttttttttacaggtagtgattattagtgcttcacatcccat 2160
QY 2161 aaaaaaacagctactaattggttaacacatatggaggtttgtgcccataatatattgcac 2220
Db 2161 aaaaaaacagctactaattggttaacacatatggaggtttgtgcccataatatattgcac 2220
QY 2221 aaaaatcatttaataataaaattattaaatcattcctcgtccattccacttgtaaat 2280
Db 2221 aaaaatcatttaataataaaattattaaatcattcctcgtccattccacttgtaaat 2280
QY 2281 gggaaattc 2288
Db 2281 gggaaattc 2288

RESULT 2
AAS76421
ID AAS76421 standard; cDNA; 1835 BP.
XX AAS76421;
AC AAS76421;
XX
DT 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #12225.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
KW
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PR
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
XX
XX DR P-PSDB; ABG12234.
DR
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 12225; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1835 BP; 533 A; 343 C; 356 G; 603 T; 0 other;

Query Match 79.1%; Score 1808.8; DB 23; Length 1835;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1810; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 477 atggggctccgtaagaagagcaccagaaccccccggttctcagccagggaattcactctg 536

Db 1 atggggctccgtaagaagagcaccagaaccccccggttctcagccagggaattcactctg 60

QY 537 cagaatcattcggaacatgctcctcgtggtgggatgtctctcgtcgtgggtgtgttc 596

Db 61 cagaatcattcggaacatgctcctcgtggtgggatgtctctcgtcgtgggtgtgttc 120

QY 597 gaggaacacagcagaacatccatcgtgtttctcactctcagcacagtggtgtgcct 656

Db 121 gaggaacacagcagaacatccatcgtgtttctcactctcagcacagtggtgtgcct 180

QY 657 gcagcagaggaacacagcccggtcctcaagtcctctattattatgtgtcacaagattg 716

Db 181 gcagcagaggaacacagcccggtcctcaagtcctctattattatgtgtcacaagattg 240

QY 717 gccacggtttcttctacatgctggtggcaatcattattcattcccaattcagggaatat 776

Db 241 gccacggtttcttctacatgctggtggcaatcattattcattcccaattcagggaatat 300

QY 777 ggttggtataaaattaaacagagaatgcagttccacaaagcgaacaaaacaaagttaac 836

Db 301 ggttggtataaaattaaacagagaatgcagttccacaaagcgaacaaaacaaagttaac 360

QY 837 gagctggtcagtttagtggtctctctttttctgtatttggggcacattcatttta 896

Db 361 gagctggtcagtttagtggtctctctttttctgtatttggggcacattcatttta 420

QY 897 atcttgaaaactgcctgcagaccaccaactcttatatggaaggtcgtccccaatagcatg 956

Db 421 atcttgaaaactgcctgcagaccaccaactcttatatggaaggtcgtccccaatagcatg 480

QY 957 atgacaatttcaaatgaagtttttctacatatccacagttcaggttactggtttatgattt 1016

Db 481 atgacaatttcaaatgaagtttttctacatatccacagttcaggttactggtttatgattt 540

QY 1017 cctgaaacttacttccagaaaaacaaaaaagacatccctcgtcaactgtctacatt 1076

Db 541 cctgaaacttacttccagaaaaacaaaaaagacatccctcgtcaactgtctacatt 600

QY 1077 ggtcttcaactcttccacattactgagcttatctctgttacttgaatcatttgggactt 1136

Db 601 ggtcttcaactcttccacattactgagcttatctctgttacttgaatcatttgggactt 660

QY 1137 cttcttttggtagctgattatttttggtagattacttccacatgtcggcgtgttttac 1196

Db 661 cttcttttggtagctgattatttttggtagattacttccacatgtcggcgtgttttac 720

Qy	1197	t	t	a	g	t	a	g	a	a	a	a	g	c	a	t	a	t	c	t	c	t	g	t	g	g	c	a	t	t	g	t	t	a	t	c	t	g	g	g	t									
Db	721	t	t	a	g	t	a	g	a	a	a	a	g	c	a	t	a	t	c	t	e	g	t	g	g	c	a	t	t	g	t	t	a	t	c	t	g	g	t											
Qy	1257	a	g	a	c	t	g	a	c	t	t	a	a	t	g	t	t	c	o	g	t	a	c	t	g	t	g	g	t	t	a	c	c	t	g	g	a	c	a	g										
Db	781	a	g	a	c	t	g	a	c	t	t	a	a	t	g	t	t	c	o	g	t	a	c	t	g	t	g	g	t	t	a	c	c	t	g	g	a	c	a	g										
Qy	1317	a	a	t	c	g	a	a	c	t	g	a	t	g	c	c	t	a	c	t	a	a	a	a	t	g	t	g	g	c	a	g	c	t	a	a	a	a	t	g	c	t								
Db	841	a	a	t	c	g	a	a	c	t	g	a	t	g	c	c	t	a	c	t	a	a	a	a	t	g	t	g	g	c	a	g	c	t	a	a	a	a	t	g	c	t								
Qy	1377	c	t	g	c	t	c	c	a	g	t	t	c	a	a	c	c	a	c	c	a	c	c	a	c	t	a	a	c	t	a	a	c	t	a	a	c	t	a	a	c	t								
Db	901	c	t	g	c	t	c	c	a	g	t	t	c	a	a	c	c	a	c	c	a	c	c	a	c	t	a	a	c	t	a	a	c	t	a	a	c	t	a	a	c	t								
Qy	1437	c	a	g	a	g	t	g	g	t	a	a	a	g	a	t	t	c	t	a	a	t	t	c	a	g	c	c	t	c	a	t	g	a	t	a	a	a	a	a	a	a								
Db	961	c	a	g	a	g	t	g	g	t	a	a	a	g	a	t	t	c	t	a	a	t	t	c	a	g	c	c	t	c	a	t	g	a	t	a	a	a	a	a	a	a								
Qy	1497	t	c	t	c	t	a	a	a	a	a	a	a	a	c	g	a	c	g	a	c	g	a	c	g	g	g	g	g	a	a	c	t	c	a	a	t	a	c	a	g	a	c	t						
Db	1021	t	c	t	c	t	a	a	a	a	a	a	a	a	c	g	a	c	g	a	c	g	a	c	g	g	g	g	g	a	a	c	t	c	a	a	t	a	c	a	g	a	c	t						
Qy	1557	c	g	c	c	a	a	a	g	a	a	a	a	a	c	t	c	a	a	t	c	t	c	a	a	t	c	t	t	g	c	a	g	c	a	t	g	a	t	g	c	t	g	c	a					
Db	1081	c	g	c	c	a	a	g	a	a	a	a	a	a	c	t	c	a	a	t	c	t	c	a	a	t	c	t	t	g	c	a	g	c	a	t	g	a	t	g	c	t	g	c	a					
Qy	1617	a	a	g	a	a	t	c	g	c	t	t	t	g	a	a	g	t	t	c	t	t	g	c	a	c	t	a	a	g	a	t	t	c	t	g	t	t	t	t	g	a	a	a	a	t	a			
Db	1141	a	a	g	a	a	t	c	g	c	t	t	t	g	a	a	g	t	t	c	t	t	g	c	a	c	t	a	a	g	a	t	t	c	t	g	t	t	t	g	t	t	t	g	a	a	a	a	t	a
Qy	1677	t	c	g	t	c	t	c	o	g	t	t	t	g	t	a	t	i	a	g	a	c	t	g	t	t	c	a	t	g	t	a	t	t	t	t	t	a	a	a	a	c	a	t	t	g	a	g		
Db	1201	t	c	g	t	c																																												

Db 1801 aaatgggaattc 1812

|||||||||||||

RESULT 3

AAK94181

ID AAK94181 standard; cDNA; 1736 BP.

XX

AC AAK94181;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human full-length cDNA, SEQ ID NO: 2724.

XX

KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX

OS Homo sapiens.

XX

PN EP1130094-A2.

XX

PD 05-SEP-2001.

XX

PF 07-JUL-2000; 2000EP-0114089.

XX

PR 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX

DR WPI; 2001-524255/58.

DR P-PSDB; AAM93265.

XX

PT 830 Primers useful for synthesizing full length cDNA clones and their

PT use in genetic manipulation -

XX

PS Claim 8; SEQ ID NO 2724; 1380pp + sequence listing; English.

XX

CC The invention relates to primers for synthesising full length cDNA

CC clones. 830 cDNA molecules encoding a human protein have been

CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA

CC molecules have been determined. Primers for synthesising the full length

CC cDNA are useful for clarifying the function of the protein encoded by

CC the cDNA. The full length clones were obtained by construction of full

CC length enriched cDNA libraries that were synthesised by the oligo-capping

CC method. The primers enable the production of the full length cDNA easily

CC without any special methods. The present sequence is a full length

CC human cDNA of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in CD-ROM format directly from EPO.

XX

SQ Sequence 1736 BP; 448 A; 358 C; 390 G; 540 T; 0 other;

Qy 2277 aaatqqqaattc 2288

```
QY 474 accatggggctccgtaagaagcaccagaagaaaccccccggtttcttaoagccaggaattcatc 533
Db 181 accatggggctccgtaagaagcaccagaagaaaccccccggtttcttaoagccaggaattcatc 240
QY 534 ctgcagaatcatcggaacatcgtctcctcgtggtgggatgtttctcctcgtcgtgggtcttg 593
Db 241 ctgcagaatcatcggaacatcgtctcctcgtggtgggatgtttctcctcgtcgtgggtcttg 300
QY 594 ttccagggaaacagcagaagcatccatcgtgtttctcaactcttcagcacagtggttgcgc 653
Db 301 ttccagggaaacagcagaagcatccatcgtgtttctcaactcttcagcacagtggttgcgc 360
QY 654 cctgcagcagagaaacagccacgggctcaaaagtcctctattattatggttcaaaagat 713
Db 361 cctgcagcagagaaacagccacgggctcaaaagtcctctattattatggttcaaaagat 420
QY 714 ttggccacaggtttctctcaatgctggtggcaattattatcattcagccaattcaggaa 773
Db 421 ttggccacaggtttctctcaatgctggtggcaattattatcattcagccaattcaggaa 480
QY 774 tatgtgttgataaaattaaacagagaaatgcagttcaccaagcgaacaaacaaagttt 833
Db 481 tatgtgttgataaaattaaacagagaaatgcagttcaccaagcgaacaaacaaagttt 540
QY 834 aacgagctcgttcagtttagtcttacttttttttttttttttttttttttttttttttttt 893
Db 541 aacgagctcgttcagtttagtcttacttttttttttttttttttttttttttttttttttt 600
QY 894 ttaactctgaaactgcgtcgaacccaactcttatataatgaaaggtcgtccccaatagc 953
Db 601 ttaactctgaaactgcgtcgaacccaactcttatataatgaaaggtcgtccccaatagc 660
QY 954 atgatgacatttcaaatgaattttctacatatcccaagttggttactggttttcatgct 1013
Db 661 atgatgacatttcaaatgaattttctacatatcccaagttggttactggttttcatgct 720
QY 1014 ttctcgaacttacttccagaacaccccaaaacaaagacatccctcgtcaacttgcctac 1073
Db 721 ttctcgaacttacttccagaacaccccaaaacaaagacatccctcgtcaacttgcctac 780
QY 1074 attggtcttcaacttccacattactgagattactctcttcttacttgaattcatttgga 1133
Db 781 attggtcttcaacttccacattactgagattactctcttcttacttgaattcatttgga 840
QY 1134 ctctctctttggtcgtcattattttgtgaattacttccacatgtcgcgcctgttt 1193
Db 841 ctctctctttggtcgtcattattttgtgaattacttccacatgtcgcgcctgttt 900
QY 1194 tactttagtgataaagtaaccagaaagcattatctctgtggccattgtgtttatcttg 1253
Db 901 tactttagtgataaagtaaccagaaagcattatctctgtggccattgtgtttatcttg 960
QY 1254 ggtagaacttgtaacttaattgttccgtactcaactgttgggtttcaacctggtcgtgatcg 1313
Db 961 ggtagaacttgtaacttaattgttccgtactcaactgttgggtttcaacctggtcgtgatcg 1020
QY 1314 cagaatcggaaatcctgagtccttactggaataatgtaaatgtgttgagcagtaaaattgct 1373
Db 1021 cagaatcggaaatcctgagtccttactggaataatgtaaatgtgttgagcagtaaaattgct 1080
QY 1374 gttctgtgtccagttgcacgataccaaagcctacgtaacatggaacttaattactctcgg 1433
Db 1081 gttctgtgtccagttgcacgataccaaagcctacgtaacatggaacttaattactctcgg 1140
QY 1434 cttccagaggtgggtgagagattcctaattcaggcctcactgtatgaaacaaacagcgtcg 1493
Db 1141 cttccagaggtgggtgagagattcctaattcaggcctcactgtatgaaacaaacagcgtcg 1200
QY 1494 agatcttctaaaaaagaacagaaacagcagtggtggagtggaacttcaaatagagtagac 1553
Db 1201 agatcttctaaaaaagaacagaaacagcagtggtggagtggaacttcaaatagagtagac 1260
QY 1554 tgtccgcaaaagagaaagaaattcttcaataattcttgcgaagcgtattgatttaattgct 1613
```

```
||||| 1261 tgtccgcaaaagagaaagagaaattctataattcttgcagcgtattgatttaattgct 1320
QY 1614 gcaaaaggaatcgtctcttgcaggtttcttctgcactagagattttctgttttgaaaa 1673
Db 1321 gcaaaaggaatcgtctcttgcaggtttcttctgcactagagattttctgttttgaaaa 1380
QY 1674 tagttcgtcctcgtcgttttcttattgaactgtttcattttttaaagacatttg 1733
Db 1381 tagttcgtcctcgtcgttttcttattgaactgtttcattttttaaagacatttg 1440
QY 1734 aggggagggaggtattattgaatgggaaagaaagattttggttgagactaaattactcat 1793
Db 1441 aggggagggaggtattattgaatgggaaagaaagattttggttgagactaaattactcat 1500
QY 1794 cgtcaaaaataatgtaaaatagttttgggtacccactatattttgtttgatttttaa 1853
Db 1501 cgtcaaaaataatgtaaaatagttttgggtacccactatattttgtttgatttttaa 1560
QY 1854 cctttcaacattttcctaattgatttcagagataaactgcacaattttgcataatgat 1913
Db 1561 cctttcaacattttcctaattgatttcagagataaactgcacaattttgcataatgat 1620
QY 1914 actggttctactccaccagttttcataataactaacaagatggtctcctctagcaaga 1973
Db 1621 actggttctactccaccagttttcataataactaacaagatggtctcctctagcaaga 1680
QY 1974 ttatgtgtttaatgcttgcgttgggttaaaaaataaaagtaacgaaaaaggtggaagtc 2029
Db 1681 ttatgtgtttaatgcttgcgttgggttaaaaaataaaagtaacgaaaaaggtggaagtc 1736

RESULT 4
AAS18576
ID AAS18576 standard; cDNA; 1517 BP.
XX
AC AAS18576;
DT 12-MAR-2002 (first entry)
DE
DE cDNA encoding human translocating chain-associated membrane protein.
KW Human; translocating chain-associated membrane protein; Biotram; ss.
XX Homo sapiens.
OS
FH Key Location/Qualifiers
CDS 153..790
FT /*tag= a
FT /product= "Translocating chain-associated membrane
protein, Biotram"
XX
PN CN1310184-A.
XX
XX 29-AUG-2001.
XX
XX 24-FEB-2000; 2000CN-0111729.
XX
XX 24-FEB-2000; 2000CN-0111729.
XX
XX (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.
PI Mao Y, Xie Y;
XX
XX WPI: 2002-034947/05.
DR P-PSDB; AAU10977.
XX
XX New human transposition chain related membrane protein and its coding
sequence -
XX
XX Claim 6; Page 18; 22pp; Chinese.
XX
XX The invention relates to a novel human translocating chain associating
```


XX WPI: 2001-202940/20.
 DR P-PSDB; AAB70695.
 XX
 PT Transformation of a cell with separate vectors expressing the sense and
 PT antisense strands of WAR-1 DNA for screening secretory and membrane
 PT proteins expressed by the cell
 XX
 PS Example 1; Page 68-71; 79pp; Japanese.
 XX
 CC The present invention describes a screening method for secretory and
 CC membrane proteins consisting of transformation of a cell with separate
 CC expression vectors for the sense and antisense RNA of DNA encoding an
 CC endoplasmic reticulum membrane protein participating in endoplasmic
 CC reticulum transport of proteins. Also described are: (1) secretory and
 CC cell membrane proteins identified by the screening method; (2) drug
 CC compositions containing these proteins; (3) host cells transformed by
 CC the separate expression vectors of the method; and (4) the preparation
 CC of secretory and cell membrane proteins by culture of the transformants.
 CC The method can be used for the identification and preparation of
 CC proteins for use in the treatment and prevention of diseases such as
 CC cancer, disorders of the nervous system, immune disorders (including
 CC allergies and rheumatism) and skeletal disorders. The present sequence
 CC encodes a specifically claimed human WAR-1 protein from the present
 CC invention.
 XX
 SQ Sequence 1288 BP; 331 A; 275 C; 276 G; 406 T; 0 other;

QY 1038 accaaaaaacaagacatccctcgtcaactgtctacatggtctcacctctccacatt 1097
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 601 accaaaaaacaagacatccctcgtcaactgtctacatggtctcacctctccacatt 660
 QY 1098 actggagcttattctctgttacttgaatcatttgggactctcttcttggtagctatt 1157
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 661 actggagcttattctctgttacttgaatcatttgggactctcttcttggtagctatt 720
 QY 1158 ttgttggaattacttccacacatgtcgccctgttttacttttagtgaataagaccag 1217
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 721 ttgttggaattacttccacacatgtcgccctgttttacttttagtgaataagaccag 780
 QY 1218 aaaggcattatctctgtggccattgtttatctctgggtagactgtgactttaattgt 1277
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 781 aaaggcattatctctgtggccattgtttatctctgggtagactgtgactttaattgt 840
 QY 1278 tcgctactcactgttgggtttccacctgggtgcacgcagaatcggaatcctgatgccct 1337
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 841 tcgctactcactgttgggtttccacctgggtgcacgcagaatcggaatcctgatgccct 900
 QY 1338 actggaatgtaaatgtgtggcagctaaattgtctgtctcgtccagttgcagatc 1397
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 901 actggaatgtaaatgtgtggcagctaaattgtctgtctcgtccagttgcagatc 960
 QY 1398 caagcctacgtacatcggaacttaattactctctgcttcagaggtgggtagaagattct 1457
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 961 caagcctacgtacatcggaacttaattactctctgcttcagaggtgggtagaagattct 1020
 QY 1458 aatattcagcctcattgtatgaaaaaagaaacggtcagatctcttaaaaaaagacagaa 1517
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1021 aatattcagcctcattgtatgaaaaaagaaacggtcagatctcttaaaaaaagacagaa 1080
 QY 1518 aacggagtggagtgaaacttcaaatagatagactgtccgcgaagaggaagagaaa 1577
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1081 aacggagtggagtgaaacttcaaatagatagactgtccgcgaagaggaagagaaa 1140
 QY 1578 tcttcaatacttctgcaagcatttaattgtctgcaaggaatctgctctcttggagt 1637
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1141 tcttcaatacttctgcaagcatttaattgtctgcaaggaatctgctctcttggagt 1200
 QY 1638 tcttctcactagagattttctgtttttgaaatagttcgtctctcttggattttgt 1697
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1201 tcttctcactagagattttctgtttttgaaatagttcgtctctcttggattttgt 1260
 QY 1698 tattaactgtttcatttttttaaa 1725
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1261 tattaactgtttcatttttttaaa 1288
 RESULT 6
 AAF74780
 ID AAF74780 standard; cDNA; 1110 BP.
 XX
 AC AAF74780;
 XX
 DT 17-MAY-2001 (first entry)
 XX
 DE Human WAR-1 encoding cDNA sequence SEQ ID NO:3.
 XX
 KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
 KW endoplasmic reticulum membrane transportation; secretory protein;
 KW cell membrane protein; cytosolic; CNS active; antiallergic; cancer;
 KW antirheumatic; nervous system disorder; immune disorder; allergy;
 KW rheumatism; skeletal disorder; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1110
 FT /*tag= a
 FT /product= "WAR-1"
 XX
 PN WO200114582-A1.

XX PD 01-MAR-2001.
XX PF 17-AUG-2000; 2000WO-JP05488.
XX PR 20-AUG-1999; 99JP-0234764.
XX PA (SUMU) SUMITOMO PHARM CO LTD.
XX PI Tohdoh N, Okuyama H, Imamura M, Ishikawa H, Nemoto K;
XX WPI; 2001-202940/20.
XX P-PSDB; AAB70695.
PT Transformation of a cell with separate vectors expressing the sense and
PT antisense strands of WAR-1 DNA for screening secretory and membrane
PT proteins expressed by the cell -
XX
PS Claim 2; Page 62-65; 79pp; Japanese.
CC The present invention describes a screening method for secretory and
CC membrane proteins consisting of transformation of a cell with separate
CC expression vectors for the sense and antisense RNA of DNA encoding an
CC endoplasmic reticulum membrane protein participating in endoplasmic
CC reticulum transport of proteins. Also described are: (1) secretory and
CC cell membrane proteins identified by the screening method; (2) drug
CC compositions containing these proteins; (3) host cells transformed by
CC the separate expression vectors of the method; and (4) the preparation
CC of secretory and cell membrane proteins by culture of the transformants.
CC The method can be used for the identification and preparation of
CC proteins for use in the treatment and prevention of diseases such as
CC cancer, disorders of the nervous system, immune disorders (including
CC allergies and rheumatism) and skeletal disorders. The present sequence
CC encodes a specifically claimed human WAR-1 protein from the present
CC invention.
XX
SQ Sequence 1110 BP; 296 A; 241 C; 236 G; 337 T; 0 other;

Query Match 48.5%; Score 1110; DB 22; Length 1110;
Best Local Similarity 100.0%; Pred. No. 6.4e-282;
Matches 1110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 atggggctccgtaagaagacagaccacgaaccccccggtcttcagcaggaattcatctg 536
DB 1 atggggctccgtaagaagacagaccacgaaccccccggtcttcagcaggaattcatctg 60
QY 537 cagaatcatcggaacatcgctctctcgctggtgggagtgtctctctgctggtggttc 596
DB 61 cagaatcatcggaacatcgctctcgctggtgggagtgtctctctgctggtggttc 120
QY 597 gaggaacagcagaaagcattccatcggtttctcactcttcagcaggtgtgctgctcct 656
DB 121 gagggaaacagcagaaagcattccatcggtttctcactcttcagcaggtgtgctgctcct 180
QY 657 gcagcagaggaaacagccacgggtccaaagtcctctattattatggtgtcaaaagatttg 716
DB 181 gcagcagaggaaacagccacgggtccaaagtcctctattattatggtgtcaaaagatttg 240
QY 717 gccacggtttcttcacatcgctggtggcaatcattatcgcacaaattcagaataat 776
DB 241 gccacggtttcttcacatcgctggtggcaatcattatcgcacaaattcagaataat 300
QY 777 ggtgtggataaaattacaagagaatcagttccaccagaagcgaacaaacaaagttaac 836
DB 301 ggtgtggataaaattacaagagaatcagttccaccagaagcgaacaaacaaagttaac 360
QY 837 gagctggtcagtttagtgttctactctttttctgtatttggggcacattcattta 896
DB 361 gagctggtcagtttagtgttctactctttttctgtatttggggcacattcattta 420
QY 897 atctctgaaactgctgtcagaccccaactcttatatggaaggctctccccaatagcatg 956

DB 421 atctctgaaaaactgctgtgtoagaccccaactcttatatggaaggctcgtccccaatagcatg 480
QY 957 atgacatttcaaatgaagttttttctacatatcccaagtgtggcttactggttttcaatgctttt 1016
DB 481 atgacatttcaaatgaagttttttctacatatcccaagtgtggcttactggttttcaatgctttt 540
QY 1017 cctgaactctacttccagaaaaaccccaaaaaacagacatccctctgccaactgtctacatt 1076
DB 541 cctgaactctacttccagaaaaaccccaaaaaacagacatccctctgccaactgtctacatt 600
QY 1077 ggtttcacctctccacattactgagagcttatctctgttacttgaaatcatttgggaatt 1136
DB 601 ggtttcacctctccacattactgagagcttatctctgttacttgaaatcatttgggaatt 660
QY 1137 ctcttttggctactgatttttggtaattactttccacatgtgcgctgtttttac 1196
DB 661 ctcttttggctactgatttttggtaattactttccacatgtgcgctgtttttac 720
QY 1197 tttagtgatgaaaaagttaccagaaaggcatatctctgtgggcatgtgtttctatcttgggt 1256
DB 721 tttagtgatgaaaaagttaccagaaaggcatatctctgtgggcatgtgtttctatcttgggt 780
QY 1257 agacttgtgaacttaattgtttccgtaactactggttttccactggttgatcgag 1316
DB 781 agacttgtgaacttaattgtttccgtaactactggttttccactggttgatcgag 840
QY 1317 aatcggaactcctgatcccttacttggaatgtaaatgtgttggaagctaaaattgctgtt 1376
DB 841 aatcggaactcctgatcccttacttggaatgtaaatgtgttggaagctaaaattgctgtt 900
QY 1377 ctgtctcagttgcagatcccaagcctcacgtaacatggaacttaattactctctggtt 1436
DB 901 ctgtctcagttgcagatcccaagcctcacgtaacatggaacttaattactctctggtt 960
QY 1437 caggggtgggttagaagattctcaattattcaggtccctcatgtatgaaaaaagacggtcgaga 1496
DB 961 caggggtgggttagaagattctcaattattcaggtccctcatgtatgaaaaaagacggtcgaga 1020
QY 1497 tcttctaaaaaagaaacagaaacggagtggtggagtggaaccttcaaatagatagactgt 1556
DB 1021 tcttctaaaaaagaaacagaaacggagtggtggagtggaaccttcaaatagatagactgt 1080
QY 1557 ccgccaagaggaagagaatcttctataa 1586
DB 1081 ccgccaagaggaagagaatcttctataa 1110
RESULT 7
AAC77810
ID AAC77810 standard; cDNA; 2833 BP.
XX
AC AAC77810;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:204.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
KW antidiabetic; antidiabetic; antirheumatic; antithrombotic; antiviral;
KW antiinflammatory; antihypertensive; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotrophic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
OS Homo sapiens.
XX
XX WO200055350-A1.
XX
PD 21-SEP-2000.

QY 1940 cataatactaaacagatggtctctctcagcaagattatgtgtttaatgcttgcgtt-ggg 1998
DB 1630 tataa---tagcagagtgcctgtctctaagaagccataatttttaagtattctttcagg 1686
QY 1999 gtaataaaagtaagaaagagtggaagtcaaa-----tcagttattctgtgta 2045
DB 1687 gtaacatggaatactataaagtgtgagtgtcaaacctttaatatgttttcagtggtctcta 1746
QY 2046 atgtgtagaattatttttaagaacttacactcagaaaaagagtgctagactcaccaaa 2105
DB 1747 attttttggaaatt-ttgttagactttacactcggaataaagattgtgaaatacccgga 1805
QY 2106 ataataaatgttctttattttacagagttagtgattatttagtcttcaccccatatt--aaa 2163
DB 1806 acaat-tgtgtgctttattttatagtgtagtggttatttagttacatccccattttaaaa 1864
QY 2164 aaaaacacagttactaaggtaacacatatagaggtgttgcctccatattattgcatcaaaa 2223
DB 1865 acaaaaacataataatggtttacaacacgtggagtttttactaacatacatattataatcaaa 1924
QY 2224 atatcat-taattataataaaaattataaaatcattcctcctgcatt 2268
DB 1925 gtatatcttcaaaagtactgttgaagtgaataatcttcttctgtgcatt 1970
RESULT 8
ID AAS62269
XX AAS62269 standard; cDNA; 2756 BP.
AC AAS62269;
DT 14-FEB-2002 (first entry)
XX cDNA sequence #56 encoding novel human secreted protein.
DE Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; gene therapy; antimicrobial; hepatotropic;
KW immunosuppressive; antirheumatic; ss.
OS Homo sapiens.
XX
XX WO200177291-A2.
XX
XX 18-OCT-2001.
XX
XX 29-MAR-2001; 2001WO-US10485.
XX
XX 06-APR-2000; 2000US-195604P.
XX
XX (GEMY) GENETICS INST INC.
XX
XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
XX Gulukota K, Graham JR;
XX
XX WPI; 2002-010900/01.
XX
XX New polynucleotides encoding secreted proteins useful for treating e.g.
XX asthma, HIV and Crohn's disease -
XX
XX Claim 1; Page 103-104; 391pp; English.
XX
XX The present invention relates to the isolation of novel cDNA sequences
XX which encode human secreted proteins. The cDNA sequences have been
XX derived from a variety of human tissues. The invention also provides
XX a method for producing proteins from these polynucleotide sequences
XX The proteins are useful for identifying compounds that modulate their
XX activity and production, and the cell is also useful for identifying
XX compounds that modulate expression of the polynucleotide sequences
XX encoding the secreted proteins. The sequences of the invention are
XX useful for treating diseases such as hyperproliferative disorders
XX (e.g. cancer), immune deficiency disorders (e.g. severe combined
XX immunodeficiency (SCID)), autoimmune disorders (e.g. multiple

CC sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory
CC disorders (e.g. arthritis) and infectious disorders (e.g. hepatitis).
CC The polynucleotide sequences of the invention are also useful in gene
CC therapy. AAS62214-AAS62838 represent the cDNA sequences of the
CC invention that encode for novel human secreted proteins.
XX
SQ Sequence 2756 BP; 825 A; 498 C; 505 G; 928 T; 0 other;

Query Match 40.4%; Score 923.4; DB 24; Length 2756;
Best Local Similarity 75.4%; Pred. No. 1.5e-232;
Matches 1392; Conservative 0; Mismatches 391; Indels 64; Gaps 17;
QY 465 ggaagcgccacatggggctcctaagaagagcaccagaagcccccggtcttcagccag 524
DB 132 ggcggcccccacatggcgattcgcaagaaagcaccagagcccccgagctgagccac 191
QY 525 gaattcatctcgcagaatcatcgcgacatcgctcctcgtcgtgggagatgttcttctgctg 584
DB 192 gaattcgtcctgcagaatcacgcggacatcgctcctcgtcgtggcgatggtcttctgctg 251
QY 585 gggcttgggttcgagggaaacagcagacatccatcggttttctcactcttcacgcacagt 644
DB 252 gggctcatgtttgagataaacggcaaaagctctctacattttgttactcttcagtcacat 311
QY 645 gttcgtgctcctgcagcagagaaacacgcccgcggctcacaagtcctctctattattatgt 704
DB 312 gtcacctccacgaacagaagaacagactgaatcagtcagtcctttattactatgac 371
QY 705 gtcaagatttggccacgggttttctctacatgctggtggcacaatattatcagccaca 764
DB 372 atcaagatttggctactgtttctctacatgctgtagggcgataattattatcgcgta 431
QY 765 attcagaataatgttggataaaaattaacaagagaatgcagttcaccaagcgaacaaa 824
DB 432 attcaagagtatagttggataaaaattaacaggcgaaatgcactctcccaacaaaacac 491
QY 825 aacaagtttaacgagctcgttcagtttagtgttctacttttttttctgttatgtgggac 884
DB 492 agcaagtttaaatgctggttcagtttagtccttacccttttgcctggttggggc 551
QY 885 acattcattttaaactctcgaataactgcctgcagaccacacatttatatgaaagctcgt 944
DB 552 acattcattctctctgaaactacatctcagaccacacatttatgagggcttat 611
QY 945 cccatagcatgatgacatttcaaatgaagtgtttctacatatccagttggcttactgg 1004
DB 612 cccataacctgatgacatttcaaatgaagtgtttctacatatcacagctggcttactgg 671
QY 1005 ttcatgcttttctgaacttacttccagaaaaaccccaaaaaaacagacatccctcgtcaa 1064
DB 672 ctccatgcttttctgaacttacttccagaaaaaccccaaaaaaacagacatccctcgtcag 731
QY 1065 ctgtctacattggtcttcacacttccacattactgagagcttatctctgttacttgaat 1124
DB 732 ctgtctacattggtcttcttactcttccacattgctgagagcttctgttacttgaat 791
QY 1125 cattgggactctcttcttgggtactgacttatttctgtgaattacttccacatgtgc 1184
DB 792 catctaggactgttctcttgggtgctacattatttggggaattcttccacattccc 851
QY 1185 ggcctgttttactttagtgtaaaagtcaccagaaaagcgcatactctgtgggccaattg 1244
DB 852 cgctgttttatttagcaatgaaagtcacagaaaagattttctctgtggcgagttctt 911
QY 1245 ttatcttggtagacttgta-ctttaattgtttccgtactcactgttgggtttcacct 1303
DB 912 ttgttttgggaagactctctgacctttaaattctctcagactgactgctgtgttggggc 971
QY 1304 ggcgtggatcgcagaaatcggaatcctgtacgttactgagaaatgtaaatgtgttggcagc 1363
DB 972 tgcaagacgacaaaatcatgaagcgtggattcagtaactggaactcaatgtgttagcgt 1031

Db 721 tttcctgaactctactttcagaaaaccccaaaacagacatncttgt-aacttggcta 779

```
Db 121 ctgtgaggagcagccaggaggagcagctgcggctgcggctgagtgatcccggaagcgcc 180
QY 474 accatggggtccgtgaagagacacaaagaccccccggttctcagcaggaattc 533
Db 181 accatggggtccgtgaagagacacaaagaccccccggttctcagcaggaattc 240
QY 534 ctgcagaatactcgacatcgctcctcgtcggtgggatgttcttctgctgggctgtg 593
Db 241 ctgcagaatactcgacatcgctcctcgtcggtgggatgttcttctgctgggctgtg 300
QY 594 ttcgagggaacagacagatccatcggtttctcactcttcagcacagtggtgtc 653
Db 301 ttcgagggaacagacagatccatcggtttctcactcttcagcacagtggtgtc 360
QY 654 cctgcagcagagacacagccagggctcaagtcctctatttatgtgtcaagat 713
Db 361 cctgcagcagagacacagccagggctcaagtcctctatttatgtgtcaagat 420
QY 714 ttggccacggtttcttctacatcgctggtggcaatcattattccacaattcaggaa 773
Db 421 ttggccacggtttcttctacatcgctggtggcaatcattattccacaattcaggaa 480
QY 774 tatgtgttgataaaattacaagagaatgcagttcaccaaaagcgaacaaacagttt 833
Db 481 tatgtgttgataaaattacaagagaatgcagttcaccaaaagcgaacaaacagttt 540
QY 834 aacgagctggtcagtttagttagtctacttttttctgtatttggggcacattcatt 893
Db 541 aacgagctggtcagtttagttagtctacttttttctgtatttggggcacattcatt 600
QY 894 ttaactctgaaactgcgtgcagaccccaactcttatatgggaagctctgcccatgc 953
Db 601 ttaactctgaaactgcgtgcagaccccaactcttatatgggaagctctgcccatgc 660
QY 954 atgatcacattcaaatgaagtgtttctacatatccc-agtggctactggtttcatgc 1012
Db 661 atgatcacattcaaatgaagtgtttctacatatccc-aatttggctactggtttcatgc 720
QY 1013 ttctcctgaactactctcagaaaaacacaaacacagacatccctcgttcaactgtcta 1072
Db 721 ttctcctgaactactctcagaaaaacacaaacacagacatccctcgttcaactgtcta 779
QY 1073 cattggtcttcacctctccaca 1095
Db 780 cattggncttcaacttttcaca 802

RESULT 12
AAZ33531
ID AAZ33531 standard; cDNA; 1120 BP.
XX
AC AAZ33531;
XX
DB 08-DEC-1999 (first entry)
XX
DE Human prostate cancer-associated EST 55.
XX
KW Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;
KW cancer; tissue specificity; human; ss.
XX
OS Homo sapiens.
XX
PN DE19811194-A1.
XX
PD 16-SEP-1999.
XX
PF 10-MAR-1998; 98DE-1011194.
XX
PR 10-MAR-1998; 98DE-1011194.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
```

```
XX WPI; 1999-519629/44.
DR P-PSDB; AAV48434.
XX
PT New nucleic acid expressed at high level in normal prostatic tissue and
PT encoded polypeptides, used to treat cancer and screen for therapeutic
XX agents -
XX
PS Claim 3; 117; 194pp; German.
XX
CC This invention describes novel nucleic acid sequences (A) that are
CC expressed at high level in normal prostatic tissue. Polypeptides (I)
CC encoded by (A) are used: (a) for identifying agents for treatment of
CC prostatic cancer and (b) for therapy of prostate cancer, optionally
CC where expressed by gene therapy methods. (A) is also used to isolate
CC full-length genes (for gene therapy) and for recombinant production of
CC (I), which can be used to raise specific antibodies. (A) are identified
CC by assembly of ESTs (expressed sequence tags) before these are analyzed
CC for expression pattern (tissue specificity). This approach eliminates
CC many of the false results, as regards tissue specificity, associated
CC with known methods that use single (usually short) ESTs. AAZ33477-233540
CC represent expressed sequence tags described in the method of the
CC invention.
XX
SQ Sequence 1120 BP; 317 A; 210 C; 206 G; 387 T; 0 other;

Query Match 29.7%; Score 679.2; DB 20; Length 1120;
Best Local Similarity 79.1%; Pred. No. 1.8e-168;
Matches 888; Conservative 0; Mismatches 213; Indels 21; Gaps 6;

QY 687 tcctctattattatggtgtcaagatttggccaggttttcttctacatgctgttgca 746
Db 1 tcctttattactatggcatcagagatttggctactgttttcttctacatgctgtgcg 60
QY 747 atcatattatcccaattcagggaatagtgttgataaaatttaacagagaatcgag 806
Db 61 ataattattcatcgtaattcaagagtatatgttgataaaatttaacagagaatcgac 120
QY 807 ttacacaaagcgaacaaacaaagtttaacagagtcgtcagtttagtgttctacttt 866
Db 121 ttctcaaaacaaacacacagcaggtttaaagtgtcagcttagtgcgttctact 180
QY 867 ttcttctgatttggggcacattcattttaaactctgaaacactgcctgcagaccact 926
Db 181 ttgctgctgttggggcacattcattctcactctctgaaacactacatctcagaccact 240
QY 927 cttatatggaagcgtcgtcccatagcatgatgacatttcaaatgaagtgtttctacata 986
Db 241 atcttatggagggttatcccatataacctgatgacatttcaaatgaagtgtttctacata 300
QY 987 tccagttggctactggtttctactgttttctcgaactctacttccagaacacacaaaa 1046
Db 301 tcacagctggctactggttctactgttttctcgaactctacttccagaacacacaaaa 360
QY 1047 caagacatccctcgtcaacattgtctacattggtcttcaacctcttccacattactggagct 1106
Db 361 gaagatattcctcgtcagctgtctacattggtcttcaacctcttccacattactggagct 420
QY 1107 tatctctgtacttgaatcaatttgggactctcttcttggtagtgcattattttgttga 1166
Db 421 tacctttgaacttgaatcaatcattagactgttcttctggtctacattattttgttga 480
QY 1167 ttactttccacatgctgcgctgttttacttttagtgatgaaagaccagaagccata 1226
Db 481 ttctttccacattcccgctgttttatttagcaatgaaagaccagaagccatt 540
QY 1227 tctctgtggccattgtgtttacttcttgggtagactgtgactttaattgttccgtactc 1286
Db 541 tctctgtggcagttcttcttgttttgggaagactctgactttaatttcttcagtactg 600
QY 1287 actgttgggttccactggtcgtggatcgagaatcgtgaatcctgactgccttactggaat 1346
Db 1287 actgttgggttccactggtcgtggatcgagaatcgtgaatcctgactgccttactggaat 1346
```

Db 601 actgtgtgtttgctgctgcaagagcagagaaatcagaaagctggtgattcagttactgactgaaac 660
 Qy 1347 gtaaatgtgtgagcgtcaaaatgctgttctgtctgctcaattgacagatccaagcttac 1406
 Db 661 ttaaatgtgtgagcgttgaatcgtgtgttctgtgctcattcgttactcagcagctt 720
 Qy 1407 gtaacatggaacttaattactctctgctcagaggtggtgtagaagattcttaatttcag 1466
 Db 721 atgatgtggaagttcatttaatttcagcttcgaaggtgaggaacattcgtctttcag 780
 Qy 1467 gcctcatgtatgaaagaaacagctc-----gagattctcaaaaaagaaca 1514
 Db 781 gcacacgtgtgaaagaaacaaacagtaactaaaggcagatctcttaaaaaaggaaca 840
 Qy 1515 gaaacacgagtg---gagtggaactcaaatagatagactgtccgccaagagga 1571
 Db 841 gaaatggtgtgaatggaacttaactcaaatgtagcagactctcccggaataaaaaa 900
 Qy 1572 gagaaatcttcaatc--tttgcagcgcattgatttaattctgcgaagggaattcgtct 1630
 Db 901 gagaaatcttcaatgaattataactaaactaattgatttaattgctcccaagaattcgtctt 960
 Qy 1631 ttgaggtttcttctgcaactgagattttctgtttttgaaat--agttcgtgtctctc 1688
 Db 961 ctactatcttctcagcattagattttctgttcttctgaaatacacagctgtgtctctt 1020
 Qy 1689 ggttttgttattgaactgtttctatgta--ttttttaagacatttgagggagagatt 1747
 Db 1021 gattttgtctattgcaggttctcatgcaattttttaaaggcattgagggagagattatt 1080
 Qy 1748 attaatgaggaataaaagattttgtgtgagactaaattac 1789
 Db 1081 gctatgaat--gaaaaaatttttagcttagactaagctac 1120

RESULT 13

AAAF74781
ID AAF74781 standard; cDNA; 1092 BP.

AC AAF74781;

DT 17-MAY-2001 (first entry)

DE Rat WAR-1 encoding cDNA sequence SEQ ID NO:4.

KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
 KW endoplasmic reticulum membrane transportation; secretory protein;
 KW cell membrane protein; cytosolic; CNS active; antiallergic; cancer;
 KW antirheumatic; nervous system disorder; immune disorder; allergy;
 KW rheumatism; skeletal disorder; ss.

OS Rattus sp.

FH Key Location/Qualifiers
 FT CDS 1..1092
 FT FT /*tag= a
 FT FT /product= "WAR-1"

PN W0200114582-AL.

PD 01-MAR-2001.

PF 17-AUG-2000; 2000WO-JP05488.

PR 20-AUG-1999; 99JP-0234764.

XX (SUMU) SUMITOMO PHARM CO LTD.

PA Tohdoh N, Okuyama H, Inamura M, Ishikawa H, Nemoto K;
 PI

XX WPI; 2001-202940/20.

DR P-PSDB; AAB70696.

XX

PT Transformation of a cell with separate vectors expressing the sense and
 PT antisense strands of WAR-1 DNA for screening secretory and membrane
 XX proteins expressed by the cell -

PS Claim 3; Page 65-68; 79pp; Japanese.

XX The present invention describes a screening method for secretory and
 CC membrane proteins consisting of transformation of a cell with separate
 CC expression vectors for the sense and antisense RNA of DNA encoding an
 CC endoplasmic reticulum membrane protein participating in endoplasmic
 CC reticulum transport of proteins. Also described are: (1) secretory and
 CC cell membrane proteins identified by the screening method; (2) drug
 CC compositions containing these proteins; (3) host cells transformed by
 CC the separate expression vectors of the method; and (4) the preparation
 CC of secretory and cell membrane proteins by culture of the transformants.
 CC The method can be used for the identification and preparation of
 CC proteins for use in the treatment and prevention of diseases such as
 CC cancer, disorders of the nervous system, immune disorders (including
 CC allergies and rheumatism) and skeletal disorders. The present sequence
 CC encodes a specifically claimed rat WAR-1 protein from the present
 CC invention.

XX Sequence 1092 BP; 260 A; 268 C; 288 G; 276 T; 0 other;

Query Match 28.3%; Score 647.8; DB 22; Length 1092;

Best Local Similarity 76.1%; Pred. No. 3.4e-160;
 Matches 84; Conservative 0; Mismatches 247; Indels 18; Gaps 3;

Qy 477 atggggctccgtaagaagcagcaccagaaacccccctgtctcagccaggaattcctctg 536

Db 1 atggggctccgcaagaagcagcaccagaaacccccctgtgtgagccaggaattcctg 60

Qy 537 cagaatcatgcggacatcgtctcgtggtgggatgtctctcctcgtgggtgtgtgttc 596

Db 61 cagaacacacgggatgtctcgtcgtgggatgtctcctcgtggacttatgttc 120

Qy 597 gagggaacagcagacgcacatcgtctcctcactctcagcagcagtggtctcctc 656

Db 121 gagggcacgcgcagagatgcagtcgtctcctcaccctgcagcagcagtggtctcc-- 178

Qy 657 gcacagagggaacagccacgggctcgaagctcctctctattattatgtgtcgaagattg 716

Db 179 -cagcggagggctaccctcgggtcccgagaccctttaccattatgggtcgaagatcgt 237

Qy 717 gccacgggtttcttctacatcgtgtggcaatcattattcattcgcacaaattcaggaat 776

Db 238 gccacagtggttcttctacatcgtgtggccatcattcattcgcacccattcagagtac 297

Qy 777 gtgttggaataataacaagagaatgcagttcaccaagcgaagcaaaaaaagtttaac 836

Db 298 gtctagataagctcagccggagactgcagtcaccaaaaggcaaaaaaattgaat 357

Qy 837 gagtctgctcagtttagtgttctactctttttttctgtatttggggcacattctatta 896

Db 358 gaggccgggcagctgagtggttctacatagtgctgtgtatctggtgtgatcatctcgt 417

Qy 897 atctctgaaaactgcctgtcagacccaactcttatatggaaggctcgtccccaatgcatg 956

Db 418 gcctctgagaactgcctgtcagacccaactctattgtggaagctcagccccaacatg 477

Qy 957 atgacatttcaaatgaaagtgtttctacatatcccaagttggcttactgttctatgtctt 1016

Db 478 atgacatttcaagtgaaattttctacatctcacaagtgtggttactgtgttctaatgttc 537

Qy 1017 cctgaactctacttccagaaaaacaaaacagacatccctcgtcactgtctacatt 1076

Db 538 ccggagctctacttccagaaagtcaggaaacaagataccccgggtcactcatctacatt 597

Qy 1077 ggttcaacctcttccacattactgagcttatctctctgtactgtgaattcattgggaatt 1136

Db 598 ggcctccacctcttccacattggaggggccttatctctgtactgtgaaccacacgtggcctg 657

XX 17-APR-2000; 2000US-0551621.
PR 08-JUN-2000; 2000US-0590751.
PR 22-JUN-2000; 2000US-0604287.
PR 20-JUL-2000; 2000US-0620405.
XX
PA (CORI-) CORIXA CORP.

PI Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepler WT;
XX
DR WPI: 2001-611721/70.

AA Breast Tumour Proteins and nucleic acids useful for the prevention,
 AT PT diagnosis and treatment of breast cancer -
 PT XX
 PS Claim 37; Page 266; 297pp; English.
 PS

The invention relates to isolated breast tumour proteins and nucleic acids that encode them, including immunogenic fragments of the proteins. Also included are expression vectors expressing the proteins, transformed cells and antibodies raised against the proteins or an antigen presenting cell expressing the protein. The proteins and nucleic acids may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate breast tumour protein expression, i.e. breast tumours and breast cancer e.g by gene therapy. The nucleic acids and their complements may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. The proteins, nucleic acids and antibodies may be used in assays to identify modulators (e.g. antagonists) of breast tumour protein expression and activity. The antibodies and antagonists may also be used to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the proteins in samples (e.g. by enzyme linked immunosorbant assay (ELISA)) and in immuno-purification diagnostic techniques. The present sequence is a cDNA from a breast tumour cDNA library isolated by subtractive hybridisation against a normal breast cDNA library.

Sequence 642 BP; 226 A; 106 C; 141 G; 169 T; 0 other;

Query Match	21.3%	Score 486.8;	DB 22;	Length 642;
Best Local Similarity	84.9%	Pred. No. 6.4e-118;		
Matches 545; Conservative	0;	Mismatches 97;	Indels 0;	Gaps 0;

Qy	641	cagtggtgctgctcctgcagcgaggagaacagccagggctccaagtccctctattatta	700
Db	642	CAATGTCACCCCTCCCAGCAACAAGAACAAGCTACTGAATCAGTGCCTCTTTATTACTA	583
Qy	701	tgggtcaaaagattggcccacggttttcttcacatgctggtggcaatcattatcatgc	760
Db	582	TGGCATCAAAAGATTGGGTACTTGTTTTCTCTCACATGCTAGTGGCGATAATATTATTCATGC	523
Qy	761	cacaaattcagggaatatggttgtgataaaaattacaagagaatgcagttcaccaaagcgaa	820
Db	522	CGTAATTCAAGAGTATATCTTGATATAAATTAAACAGGCGAATGACACTCTCCAAAAAANA	463
Qy	821	acaaaaaacagtttaacagdcctggtgcagtttagtgtgtctactttttttctgtatttg	880
Db	462	ACACAGCAAAGTTTAATGAATCTGGTCAGCTTAGTCGGTCTACCTTTTTTGCTGTGTTTG	403
Qy	881	ggcaccaticattttaacctctgaacctgcctgtcagacccaactcttatatgggaaggc	940
Db	402	GGGCACATCATTTCTCATCTCGAAACATACATCTCAGACCACCACTATCTTTATGGAGGGC	343
Qy	941	tgcgtccccatagcatgatgacatttcacaaatgaagttttctcacatatccoaegtggctta	1000
Db	342	TTATCCCCATAACCTGATGACATTTCAAATGAAGTTTTTCTCATATACACAGCTGGCTTA	283
Qy	1001	ctggtttctatgcttttctcgtgaactactcttcagaaaaaccaaaaaaacagacatccctcg	1060
Db	282	CTGGCTTCATGTTTTGTTTTCCCTGAATCTACTTCCAGAAAACCAAAAAGAAATATCTCTCG	223

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 15:16:46 ; Search time 97.09 Seconds
(without alignments)
5788.541 Million cell updates/sec

Title: US-09-807-470-3

Perfect score: 2288

Sequence: 1 tataggcacgcgtggtcga.....ccactgtaaatggaattc 2288

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58.2	2.5	7218	1 US-08-232-463-14	Sequence 14, Appl
2	44	1.9	1431	4 US-09-316-083-2	Sequence 2, Appli
3	41.2	1.8	5506	4 US-09-004-838-93	Sequence 93, Appl
4	40.8	1.8	9636	1 US-08-323-170B-1	Sequence 1, Appli
5	40.8	1.8	9636	4 US-08-954-441-1	Sequence 1, Appli
6	39.8	1.7	549	4 US-09-342-681C-6	Sequence 6, Appli
7	38.8	1.7	8495	4 US-09-797-906-3	Sequence 3, Appli
8	39.6	1.7	19124	2 US-08-487-826B-13	Sequence 13, Appl
9	39.4	1.7	2791	4 US-09-570-367C-1	Sequence 1, Appli
10	39.2	1.7	8585	1 US-08-030-096-3	Sequence 3, Appli
11	38.6	1.7	767	4 US-08-998-416-472	Sequence 472, App
12	38.2	1.7	583	4 US-09-123-912-100	Sequence 100, App
13	38	1.7	2301	4 US-09-561-825-1	Sequence 1, Appli
14	38	1.7	2361	4 US-09-561-825-26	Sequence 26, Appl
15	38	1.7	2362	4 US-09-561-825-27	Sequence 27, Appl
16	38	1.7	2363	4 US-09-561-825-28	Sequence 28, Appl
17	37.6	1.6	1018	1 US-08-444-083-6	Sequence 6, Appli
18	37.6	1.6	1018	1 US-08-286-304-6	Sequence 6, Appli
19	37.6	1.6	1018	1 US-08-442-745-6	Sequence 6, Appli
20	37.6	1.6	1018	1 US-08-443-129-6	Sequence 6, Appli
21	37.6	1.6	1018	1 US-08-443-952-6	Sequence 6, Appli
22	37.6	1.6	1018	1 US-08-443-130-6	Sequence 6, Appli
23	37.6	1.6	1018	3 US-08-898-911-6	Sequence 6, Appli
24	37.6	1.6	1018	5 PCT-US95-04467-6	Sequence 6, Appli
25	37	1.6	921	4 US-09-377-648-4	Sequence 4, Appli
26	36.8	1.6	19124	2 US-08-487-826B-13	Sequence 13, Appl
27	36.4	1.6	2361	4 US-09-561-825-29	Sequence 29, Appl

28 36.4 1.6 4403765 4 US-09-103-840A-2 Sequence 2, Appli
29 36.2 1.6 3750 3 US-08-617-860B-19 Sequence 19, Appl
30 35.8 1.6 782 4 US-09-007-119-15 Sequence 15, Appl
31 35.8 1.6 3850 3 US-08-329-799-34 Sequence 34, Appl
32 35.6 1.6 860 4 US-08-998-416-287 Sequence 287, App
33 35.6 1.6 7218 1 US-08-232-463-14 Sequence 14, Appl
34 35.4 1.5 6216 3 US-09-213-053-1 Sequence 1, Appli
35 35.2 1.5 663 4 US-08-998-416-131 Sequence 191, App
36 35.2 1.5 711 4 US-08-998-416-786 Sequence 786, App
37 35.2 1.5 724 4 US-08-998-416-683 Sequence 683, App
38 35.2 1.5 732 4 US-08-998-416-1036 Sequence 1036, Ap
39 35.2 1.5 828 4 US-08-998-416-538 Sequence 538, App
40 35.2 1.5 854 4 US-08-998-416-534 Sequence 534, App
41 35.2 1.5 860 4 US-08-998-416-287 Sequence 287, App
42 35.2 1.5 1736 3 US-09-182-816-22 Sequence 22, Appl
43 35.2 1.5 1736 3 US-09-182-816-22 Sequence 24, Appl
44 35.2 1.5 1736 3 US-09-471-528-22 Sequence 22, Appl
45 35.2 1.5 1736 3 US-09-471-528-24 Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

```
Query Match          2.5%; Score 58.2; DB 1; Length 7218;
Best Local Similarity 5.2%; Pred. NO. 1.7e-05;
Matches 21; Conservative 221; Mismatches 159; Indels 0; Gaps 0;

Qy 849 tttagtggtctacttttttctgtatttggggacatcattttaaactcgaac 908
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1073 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1132

Qy 909 tgcctgcagaccacactttatatgaagctcgtcccatagcatgatattca 968
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1133 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1192

Qy 969 atgaagttttcacatccagcttggtctactggttctcatgctttctcgaactac 1028
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1193 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1252

Qy 1029 ttccagaaaaaacacacacacccctcgtcaactgtctacattggtctcaactc 1088
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1253 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1312

Qy 1089 ttccacattactgagcttatctctgtacttgatcatcttggaactctcttttgta 1148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1313 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1372

Qy 1149 ctgcattatttggtaacttctccacatgctgcgcctgtttacttttagtgtaa 1208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1373 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1432

Qy 1209 aagtaccagaaagcatatctctgtggccattgtgttat 1249
   ||||| | ||||| | ||| | |
Db 1433 YYGACCAAAATCTCTACTCTCTTAACTACTTGCATAGAT 1473

RESULT 2
US-09-316-083-2
; Sequence 2, Application US/09316083A
; Patent No. 6280942
; GENERAL INFORMATION:
; APPLICANT: The Institute of Physical and Chemical Research
; TITLE OF INVENTION: Endonuclease
; FILE REFERENCE: PH-651
; CURRENT APPLICATION NUMBER: US/09/316,083A
; CURRENT FILING DATE: 1999-05-20
; EARLIER APPLICATION NUMBER: JP98/141861
; EARLIER FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1431
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1428)
US-09-316-083-2
```

```
Query Match          1.9%; Score 44; DB 4; Length 1431;
Best Local Similarity 53.8%; Pred. NO. 0.044;
Matches 112; Conservative 0; Mismatches 95; Indels 1; Gaps 1;

Qy 2038 attctgtaattgttagaattattttttaagaacttcacacactcagaaagattgctagac 2097
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 285 attattatactatttttattatttttaataataaataaataaataaataaataaataa 344

Qy 2098 tcacaaaaataataaattgtttttattttacaggtagtgattattagcttcaatcccca 2157
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 345 taataataataataataattttcattataataataataataataataataataataa 404

Qy 2158 tttaaaaaaacacagtagtactaatgggttaacacatatgagggtttgtctgccatataatgc 2217
   ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db 405 tttaaatataaaactcttaattgaa-actaatattatctataaatttttttaattg 463
```

```
Qy 2218 atcaaaataatcattaaataataataaaaa 2245
   || | | | | | | | | | | | | | | | |
Db 464 ataaataatttaataataataataatgaa 491

RESULT 3
US-09-004-838-93/c
; Sequence 93, Application US/09004838
; Patent No. 6350933
; GENERAL INFORMATION:
; APPLICANT: Michelmore, Richard W.
; APPLICANT: Shen, Kathy
; APPLICANT: Meyers, Blake
; TITLE OF INVENTION: Procedures and Materials for
; CONFERRING PEST RESISTANCE IN PLANTS
; NUMBER OF SEQUENCES: 140
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/004,838
; FILING DATE: 09-JAN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/781,734
; FILING DATE: 10-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-078810DS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5506 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..5506
; OTHER INFORMATION: /note= "RG2D"
US-09-004-838-93
```

```
Query Match          1.8%; Score 41.2; DB 4; Length 5506;
Best Local Similarity 44.1%; Pred. NO. 0.49;
Matches 210; Conservative 1; Mismatches 264; Indels 1; Gaps 1;

Qy 1812 atagttttggggatccaccactatattttgttgatttttaaacctttcaacattttccta 1871
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5150 AGAATTTTCTTGATCATAGTAAATTTAGTTCGCTTTTATTTTAAATGTTGT-CTT 5092

Qy 1872 atgatttcgagagataactgcacaaattttgcataatgatactggtttactctccac 1931
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5091 TTTTATTATGTTTCTTCTTAATTCGTTTATTTTATTTACCGTGATCTTTTATTG 5032

Qy 1932 cagtgtttcataactacaagaatggtctcctcagcaagattatgttttaagtctg 1991
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5031 TATTAGTGCCTTTTAAATAAAAAAATGTTATGTTTATTTTAAATTTTCTAAATTA 4972

Qy 1992 cttgggggttaaaataaaagtcacgaaaaagggtggagtcacaaatcagtagtattctgttaattgtt 2051
```

Db 4971 ATTAAATTAATAATGATGTGGTAGTGACTTAAATGTGTTGGAATAATGCTTATTTT 4912
QY 2052 agaatttatttttaagaactcagaagaagctagactcaccacaaataa 2111
Db 4911 ATGTGTTAGAGTTGAAAAGAGTGAGGCAAGAGAGAGAGAGAAACGGATATA 4852
QY 2112 aatgtctttttttcacaggtagtgattattagtgcttccatcccccatttaaaaaaacaca 2171
Db 4851 AACTGTGTGACAAATGAGTTAAGTGAGATTGACTAGTTCCTACCACTTTAAAA 4792
QY 2172 gtactaatgggtaacacataggaggttgctgcccataatatgtgcatcaaaatcatc 2231
Db 4791 ATCATTTACATATACAAATATACATGACTACAGAAAAAATAATTTTATTATATATCGCT 4732
QY 2232 aattaataaaaaattaaaaatcattcctgctccatccacttgtaaatgggaatt 2287
Db 4731 TTGTAATACAAATTAATAAAGTGTAACTGTTAATAACATATCCAATTTCAATTTGTAAT 4676
RESULT 4
US-08-323-170B-1
; Sequence 1, Application US/08323170B
; Patent No. 5733772
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; APPLICANT: Kaslow, David C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pfs230
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,170B
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,409
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Quine, Jonathan A.
; REGISTRATION NUMBER: P-41,261
; REFERENCE/DOCKET NUMBER: 015280-113100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 149..9556
US-08-323-170B-1
Query Match 1.8%; Score 40.8; DB 1; Length 9636;
Best Local Similarity 49.3%; Pred. No. 0.83;
Matches 134; Conservative 0; Mismatches 137; Indels 1; Gaps 1;

QY 1856 tttaaacatttcttaataatgatttgcagagataaactgcacaaatttgcataatcagatac 1915
Db 3000 TTGATACATTACCTTAAATGTAATGTTAAATAAAACACAATATCCAAATATCGAGATAT 3059
QY 1916 tggttcttactccaccagtggtttcataataactaaacaagatggctctctctctagcaagatt 1975
Db 3060 TTCTCTAAAACATTTAAAGATATAAAGGAAGTATTAAATTAGATCTTGATATACAAATATC 3119
QY 1976 atgtgttttaagtctgtcttgggggttaataaaaaagtlacg-aaaaagggtggaagctcaaatc 2034
Db 3120 AATGTTTATAGTAAATTTTAAATTCATACACAGAAATGCAAAAGTATTTTAAATTTATATC 3179
QY 2035 agtattctgttaattgttagaatttttttaagaacttacaactcagaaaaagattgcta 2094
Db 3180 CATATTATTATTTTTCATTTTAAATCATATAGAAAAAGAAATTAATAAAATAATCCCTA 3239
QY 2095 gactcccccataataaatgttcttattttt 2126
Db 3240 CATATAAAATCATAAAGATGTGAAATATTTT 3271
RESULT 5
US-08-954-441-1
; Sequence 1, Application US/08954441
; Patent No. 6316000
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; APPLICANT: Kaslow, David C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pfs230
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/954,441
; FILING DATE: 20-OCT-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/323,170
; FILING DATE: 13-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,409
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 015280-113110US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 149..9556
US-08-954-441-1

US-08-030-096-3

Query Match 1.7%; Score 39.2; DB 1; Length 8585;
Best Local Similarity 50.0%; Pred. No. 2.1;
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1713 tgaatttttaagacattgagggagggagggattattatgaatgggaaaaaagatttt 1772
DB 6791 TTTGCTTCACATACAAATGATTTAAACGACATTTTTCATAATATGGATGACATAGT 6850

QY 1773 gtttgagacaaattactcatcgtaacaaataatgtcacaataatgttttggggatccaccact 1832
DB 6851 TCTTACGTAATTTATTTTGATATTTTAACTAATTTATATATTTTATGCTTTTCCRAAT 6910

QY 1833 atattttgttgatttttaacatttcaacatttccctaataatgatttgcagagataactgc 1892
DB 6911 ACATACAGTAGTGTGTCCTTAAATAATCAAAATTTTATACGTAATGTTTAAATATATGCAC 6970

QY 1893 acaattttgcatatca 1908
DB 6971 ACAATTCCTTAAAAACCA 6986

RESULT 11

US-08-998-416-472/c
; Sequence 472, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 472:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 767 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:

; ORGANISM: PAG1337RP
US-08-998-416-472

Query Match 1.7%; Score 38.6; DB 4; Length 767;
Best Local Similarity 49.3%; Pred. No. 0.87;
Matches 101; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 2054 aattttttttaagaacttcacaactcagaaaagattgctagactccaccaataataaaa 2113
DB 570 AATTATATACCTTTAGATAATAAATATTAGATTTAACTATCTTAATAATATATTATT 511

QY 2114 tgtttttattttacaggtagtgattattagtgcttccatcccccatttaaaaaaacagt 2173
DB 510 AGGTAATAATTTAGTAGGTAGTAATATCCCAATTAAGGGGAGACTATTAAATAGAAATATT 451

QY 2174 actaatgggtaacacatatgggggttgcgcataatatatgcatacaataatcatctaa 2233
DB 450 ACTAGACTAATAAATAAATAATATTATGAAGGTACATTTTAATAATATATATATCAATGA 391

QY 2234 ttaataaaaaatattaaaaatcatt 2258
DB 390 AGTAAATTAATAATATTATATAAT 366

RESULT 12

US-09-123-912-100
; Sequence 100, Application US/09123912A
; Patent No. 6312695
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455C1
; CURRENT APPLICATION NUMBER: US/09/123,912A
; CURRENT FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: 09/040,802
; PRIOR FILING DATE: 1998-03-18
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 100
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (404)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (506)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (514)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (527)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (528)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (538)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (548)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (556)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (568)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base


```

; ORGANISM: Glycine max cv. Williams 82
;
; FEATURE:
;
; NAME/KEY: promoter
; LOCATION: (1)..(2362)
US-09-561-825-27

```

	Query Match	1.7%	Score 38;	DB 4;	Length 2362;
	Best Local Similarity	50.0%	Prod. No. 2.2;	Mismatches	95; Indels
	Matches	95;	Conservative	0;	Gaps
QY	1698	tattgaactgtttccatgtattttttaagacatttgaggaggaggagatttattatgaatg	1757		
Db	955	TATCGAAATGATTATTTCGTAAAGGAAAGAAATCTTTTAAAGTAGTCCTATTCTCTAAAT	896		
QY	1758	ggaaaaaagaattttgggttgagactaaataactcgtccaaaaataatgtccaaatagtt	1817		
Db	895	AAAAAGTAAAAATGATTATTTGCTTTTAATTTAATATTTATCGAACATTTCTATTGAAAA	836		
QY	1818	ttggggatccaccactatatttgttttgatttttaacctttcaacattttcccaatgatt	1877		
Db	835	TTTAAGATAAATATTATTATTTTAAAGATAAATGAATTTTCATAAATGATCTATTATA	776		
QY	1878	tgcagagata	1887		
Db	775	TGAAAAAAA	766		

Search completed: September 6, 2002, 15:18:39
Job time: 7823 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 13:07:54 ; Search time 3580.18 Seconds
(without alignments)
8625.553 Million cell updates/sec

Title: US-09-807-470-3

Perfect score: 2288

Sequence: 1 tataggcgacgggtgtcg.....ccactgtaaatgggaattc 2288

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estopl:*

7: em_estro:*

8: em_htc:*

9: gb_estl:*

10: gb_est2:*

11: gb_htc:*

12: gb_gss:*

13: em_gss_hum:*

14: em_gss_inv:*

15: em_gss_pln:*

16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	785.8	34.3	1356	11 BC020283	BC020283 Homo sapi
2	698.2	30.5	760	10 BG714759	BG714759 602673845
3	692.6	30.3	1064	9 AL523342	AL523342 AL523342
4	608.2	26.6	1142	10 BM466240	BM466240 AGENCOURT
5	577.6	25.2	1135	10 BM466239	BM466239 AGENCOURT
6	555.2	24.3	822	9 AU130188	AU130188 AU130188
7	553.2	24.2	814	9 AU132871	AU132871 AU132871
8	532.2	23.3	779	10 BM045930	BM045930 603624955
9	511.4	22.4	790	10 BE612552	BE612552 601452065
10	509.4	22.3	879	10 BI870787	BI870787 603394607
11	508.2	22.2	905	9 AU131342	AU131342 AU131342
12	503.8	22.0	727	10 BG547017	BG547017 602573813
13	500.6	21.9	836	10 BF969571	BF969571 602271922
14	497.4	21.7	879	9 AU128453	AU128453 AU128453
15	495.2	21.6	910	10 BM459490	BM459490 AGENCOURT
16	483.6	21.1	944	10 BI655670	BI655670 603283957
17	483	21.1	483	9 AV725026	AV725026 AV725026

18	481.8	21.1	769	10 BG542329	BG542329 602571885
19	481.8	21.1	808	10 BI830439	BI830439 603073495
20	476	20.8	496	9 AW078666	AW078666 xb02907.x
21	472.2	20.6	660	10 BF984410	BF984410 602307936
22	468	20.5	969	10 BG975574	BG975574 602845246
23	462.8	20.2	1015	10 BG389557	BG389557 602414256
24	462	20.2	878	10 BE618241	BE618241 601462496
25	459	20.1	695	10 BI733351	BI733351 603353253
26	453.6	19.8	651	10 BG714581	BG714581 602671468
27	453	19.8	476	9 AI138349	AI138349 qd87a01.x
28	451.8	19.7	859	10 BG323904	BG323904 602422933
29	447.2	19.5	636	10 BG077352	BG077352 H3014E03-
30	445	19.4	787	10 BE788046	BE788046 601482653
31	444	19.4	450	9 AW023347	AW023347 df53f06.y
32	444	19.4	608	10 BG614318	BG614318 602643420
33	444	19.4	614	9 AL602739	AL602739 DKZP6866F
34	442.8	19.4	1067	10 BF538214	BF538214 602053686
35	436.8	19.1	944	10 BG916945	BG916945 602816230
36	433.6	19.0	504	12 AQ127602	AQ127602 HS_3061_A
37	432.6	18.9	493	9 AA169554	AA169554 zp17b09.s
38	432	18.9	835	10 BE887450	BE887450 601509592
39	426.6	18.6	569	9 AI201635	AI201635 qb81d09.x
40	424.2	18.5	825	10 BI731069	BI731069 603351703
41	423.8	18.5	736	10 BF141338	BF141338 601789863
42	422	18.4	1048	9 AL523341	AL523341 AL523341
43	421.4	18.4	708	10 BE914578	BE914578 601665679
44	418.4	18.3	773	10 BI730813	BI730813 603351811
45	417.6	18.3	731	10 BG700572	BG700572 602680806

ALIGNMENTS

RESULT 1

BC020283 LOCUS Homo sapiens, clone IMAGE:4519858, mRNA. 1356 bp linear HTC 20-DEC-2001

ACCESSION BC020283.1 GI:17946782

VERSION BC020283.1

KEYWORDS HTC.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1356)

Strausberg,R.

Direct Submission

Submitted (20-DEC-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs-remail.nih.gov

Tissue Procurement: DCTD/btp

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: B#CM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: villalon@bcm.tmc.edu

Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,

A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,

Muzny, D.M., Gunaratne, P., Yoon, V., Kowis, C., Martin, R.,

Lawrence, S., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 29 Row: h Column: 1

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis


```
RESULT 5
LOCUS BM466239 1135 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6457038 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5576991
5', mRNA sequence.
ACCESSION BM466239
VERSION BM466239.1 GI:18515281
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1135)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM12329 row: 1 column: 16
High quality sequence stop: 710.
FEATURES
source
location/Qualifiers
1..1135
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5576991"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
1 others
BASE COUNT 285 a 259 c 241 g 349 t
ORIGIN
Query Match 25.2%; Score 577.6; DB 10; Length 1135;
Best Local Similarity 82.9%; Pred. No. 9.2e-108;
Matches 695; Conservative 0; Mismatches 139; Indels 4; Gaps 3;
QY 465 ggaagcgccaccatgggctccgtaagaagcagcaccagaacccccccgttctcagccag 524
DB 41 GCGCGCCCCACCATGCGATTTCGCAAGAAAGCACCAGAGCCCCCAGTGTGAGCCAC 100
QY 525 gaattcactctgcagaatactgcgacatcgctctcctcgtggtgggatgttcttctgctg 584
DB 101 GAATTCGCTCGCTCAGAAATCAGCGGACATCGTCTGCTGTGGCGATGCTCTCTGCTG 160
QY 585 gggctgtgttcagggaacagcagaacatccatcggtttctcactcttcacagcagcagt 644
DB 161 GGGCTCATGTTTGAGATACGGCAAAAGCTTCTATCATTTTTTGTACTCTTCAGTCAAT 220
QY 645 gttgctgtcctgcgacgaggaacaaagccagggctcaaaatccctctattattatggt 704
DB 221 GTCAACCTCCACGACAGACAGAAACAGCTACTGAATCAGTGTCTCTTATTACTATGCG 280
QY 705 gtaagaagtgtggccacggttttctctacatgctggtgggaatcattattatccacaca 764
DB 281 ATCAAGAGATTTGGCTACTGTTTCTCTACATGCTAGTGGCGATAATATTATCCGCGTA 340
QY 765 attcaggaatgtgttgataaaaattacaagaagaatgcagttcaccaagaagcgaacaa 824
DB 341 ATTCAGAGATATATGTTGGATAAAAATTAACAGGCGGAATGCACTTCTCCAAACAAACAC 400
QY 825 aacaagttaacagcagtcgtggtcagtttagtgtgttctacttttttcttctgtattggggc 884
Db 401 AGCAAGTTTAAAGTAATCTGGTCAGCTTAGTGGCTTCTACTCTTTTGCCTGTCTTGGGCG 460
QY 885 acattcatttbaatctctgaaacacgctgctcagaccccaactcttatatggaagcctgt 944
Db 461 ACATTCAATTCATCTCTGAAACATACATCTCAGAGCCCAACATCTATTATGAGGGCTTAT 520
QY 945 ccccatagcatgatcacatttcaaatgaagtttttctacatctccagttggtcttactg 1004
Db 521 CCCCATAACTGATGACATTTCAATGAAGTTTCTTACATATCACAGCTGGCTTACTGG 580
QY 1005 ttctatgcttttccctgaactctacttccagaaaccccaaaacagacacccctcgtcca 1064
Db 581 CTTTCATGCTTTTCTGAACTCTACTTCCAGAAACCAAAAGAGATATTCTCTGTCAG 640
QY 1065 ctgtctacatggttcttcacactcttcacatctactgagcttactcttcttactgta 1124
Db 641 CTTGTCTACATTTGGCTTTTACCTCTTCCACATTTGCTGGAGCTTACCTTTTGAACCTGAAT 700
QY 1125 catttggactctcttcttggtagctgactgatttttggtaatttcttcttccacatgtgc 1184
Db 701 CATCTAGGACTTGTCTTCTGCTGTCTACATTTTGTGTAATTTCTTTTCCACATTTTC 760
QY 1185 ggcctgttttacttttagtgat-gaaaagtaccagaaaggca-tatctctgtgggcca--t 1240
Db 761 CGCCTGTTTATTATTAGCAATGGAAGATATCAGAAAGGGAATTCTCTGTGGCCAGNT 820
QY 1241 tgtgtttatctgggtagactgtgactttaattgtttccgtactactcactgttggggttt 1298
Db 821 CTTTGTGTTTTTGGGAAAACCTTCTGACTTTAATCTCTTTCAGAACTTGACTGGGTGTTT 878
RESULT 6
LOCUS AU130188 822 bp mRNA linear EST 24-OCT-2000
DEFINITION AU130188 NT2RP3 Homo sapiens cDNA clone NT2RP3000388 5', mRNA
sequence.
ACCESSION AU130188
VERSION AU130188.1 GI:10990542
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 822)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
location/Qualifiers
1..822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP3000388"
/clone_lib="NT2RP3"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/notes="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 213 a 171 c 148 g 287 t
ORIGIN
```

```
Query Match      24.3%; Score 555.2; DB 9; Length 822;
Best Local Similarity 82.0%; Pred. No. 3.4e-103;
Matches 638; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 622 tgttttcactcttcagcacagtggttgcgtccctgcagcagagagaacagcccggtct 681
Db 1 TTTTGTGTACTCTTCAGTACAATGTCACCCCTCCAGCAACAGAGAACAAGTACTGAAT 60

Qy 682 caaagtcctctattattgtgtcaaaagattggccacggttttcttcacatctgg 741
Db 61 CAGTGTCCCTTTATCTACTGGCATCAAGATTGGCTACTGTTTCTTCTACATGCTAG 120

Qy 742 tggcaatcattatcctgcacaaattcaggaaatgtgttgatgataaaattaaacagagaa 801
Db 121 TGGCGATAATTTATTCATGCGCGTAATCAAGAGATATATGTTGGATAAAATTAACAGGCGAA 180

Qy 802 tgcagttccacaaagcgaaacaaacaaagtttaacagagtcgtgcagtttagtgttct 861
Db 181 TGCACCTTCTCCAAACAAACACACAGCAAGTTTAATGAATCTGGTCAGTTAGTGCCTTCT 240

Qy 862 actttttctgtatttggggcacatttcatttaactctctgaaactgcgtgcagacc 921
Db 241 ACCTTTGTGCGTGTGTTGGGCGACATTCATCTCATCTCTGAAACCTACATCTCAGACC 300

Qy 922 caactttatatgaagagctgctcccatagcatgatgacatttcaaatgaagttttct 981
Db 301 CAACATCTTATGAGGGCTTATCCCCATACCTGATGACATTTCAAATGAAGTTTCTCT 360

Qy 982 acatccagctgtggttactggttgcatttgcatttgcatttgcatttgcatttgcatttgc 1041
Db 361 ACATATCACAGCTGGCTTACTGGCTTCATGCTTTTCTGAACTTACTTTCAGAAAACCA 420

Qy 1042 aaaaacagaacacctcgtcaactgtctacattggtcttcacactcttcacacttactg 1101
Db 421 AAAAAGAAGATATTCCTCGTCAGCTGTGTCATATGTCCTTACTCTTCCACATTTGCTG 480

Qy 1102 gagcttatctgttacttgatcatttgggaacttctcttggctactgctgatttttg 1161
Db 481 GAGCTTACCTTTTGAACCTTGAATCATCTAGGACTTGTCTCTCATGTCCTCATTTATTTG 540

Qy 1162 ttgaattacttcccacatgctgcgctgttttacttacttagtgaagaagttaccagaag 1221
Db 541 TTGAATTTCTTTTCCACATTTCCGCGCTGTTTATTTATTTAGCAATGAAAGTATCAGAAAG 600

Qy 1222 gcatactctgtggccatgtgtttacttctgttgtagactgtgaactttaaattttccg 1281
Db 601 GATTTTCTCTGGGCGAGTCTTTTGTGTTGGGAAGACTTCTGACTTAAATCTTTTCAG 660

Qy 1282 tactcaactgttgggtttcacctggtggtgatcgagaatcggaaatcctgatgaccttactg 1341
Db 661 TACTGACTGTGTTTGGCCCTTCAGAGAGCAGAAAATCAGAACTGGATTTCAGTACTG 720

Qy 1342 gaaatgtaaatgtgtggcagctaaaattgctgtctgtccagttccagttgacgatcca 1399
Db 721 GAAACTTCAATGNGTTAGCTGCTGAGATCGCTGCTGGCATCATTTGCGTTACTCA 778
```

```
RESULT 7
AU132871 AU132871 NT2RP4 Homo sapiens 814 bp mRNA linear EST 24-OCT-2000
LOCUS AU132871 NT2RP4 Homo sapiens cDNA clone NT2RP4000757 5', mRNA
DEFINITION sequence.
ACCESSION AU132871
VERSION AU132871.1 GI:109933410
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 814)
AUTHORS Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J.,
```

Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuho,Y. and Isogai,T.
HRI human cDNA project (Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuho,Y., Isogai,T.)
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
Location/Qualifiers
source
1. 814
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP4000757"
/clone_lib="NT2RP4"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/notes="Vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 206 a 195 c 161 g 249 t 3 others
ORIGIN

Query Match 24.2%; Score 553.2; DB 9; Length 814;
Best Local Similarity 82.4%; Pred. No. 8.8e-103;
Matches 645; Conservative 0; Mismatches 136; Indels 2; Gaps 1;
Qy 465 ggaagcgccaccatggggctccgttaagaagagcaccagaaccccccgcttcctcagcag 524
Db 33 GGGCGCCCCACCATGGCGGATTTCGCAAGAAAAGCACCAGAGCCCCCGAGTGTGAGCCAC 92
Qy 525 gaatttcactctcagaatcatcgagacatgctctccctgcgtgggagtgcttctcctcgt 584
Db 93 GAAATTCGCTTCGAGAATACAGCGGACATGCTCTCTGTGTGGCGATGGTCTTCTCTGTG 152
Qy 585 gggctgtgttcagggaacagcagagcatccatcgtgtttctcactcttcacacagct 644
Db 153 GGGCTCATGTTGAGATAACGGCAAAAGCTCTATCATTTTGTGTACTCTTCAGTACAAT 212
Qy 645 gttgctgcccctgcagcagaggaacagccacgggctcaagtcctcctctattatgg 704
Db 213 GTCACCTCCCGACCAACAGAGAACAAGTACTGAATCAGTGTCCCTTATTTACTATGGC 272
Qy 705 gtcaagatitggccacggttttctctacatgctgtggcaatcattatctatctgcaca 764
Db 273 ATCAAGAATTTGGCTACTGTTTCTCTACATGCTAGTGGGATPAATATTTCATGCCGTA 332
Qy 765 attcaggaatatgtgttggataaaattaaacagagaatgcagttccaccaagcgaacaa 824
Db 333 ATTCAAGAGTATATGTTGGATAAAATTAACAGGCGAATGCATCTTCCAAAACAAACAC 392
Qy 825 acaagtttaacagagctgtgctcagtttagtgttctcacttttttcttatttggggc 884
Db 393 AGCAAGTTTAATGAATCTGCTAGCTTAGTGGCTTCTACCTTTTGGCTGTGTGGGGC 452
Qy 885 acatttcatttaactctctgaaactgcctgcagacccaactcttatatgaaggtcgt 944
Db 453 ACATTCATTTCTACTCTGAAACTACATCTCAGACCCCACTATCTTATGGAGGCTTAT 512
Qy 945 ccccatgacatgatgacatttcaaatgaagtgttttctacatatccccagttggttactgg 1004
Db 513 CCCCATAACTGATGACATTTCAAATCAAGTTTCTTACATATACACAGCTGGCTTACTGG 572
Qy 1005 ttcatgcttttctcgtgaactctacttccagaaaaacacacacacacacacacacacac 1064

Db 573 CTTTCATGCTTTCTCCTGAACCTCTACTTCCAGAAAAACCAAAAAAGAGATATTCCTCGTCAG 632
 QY 1065 cttgtctacattggtttctcaacctctccacattactcggagcttatctctgttacttgaat 1124
 Db 633 CTTGTCTACATTGGTCTTACCTCTCCCATTTGCTGGAGCTTACCTTTTGAACCTTGAAT 692
 QY 1125 catttgggacctctctcttcttggctactgcattatttcttgaattactcttccacatgtgc 1184
 Db 693 CATCTAGGACTTGGCTCTCTGGTGCTACATTATTT--GTGAATTTCTTTTTCACATTTC 750
 QY 1185 ggcctgtttactttagtgcataaaagaccagaaaggacatactctgtgggcattgtg 1244
 Db 751 CGCCTGCTTATTATTAGCANTGAAAAGTATCAGAAANGATTTCTCTGGGGCANGTCT 810
 QY 1245 ttt 1247
 Db 811 TTT 813
 RESULT 8
 LOCUS BM045930 779 bp mRNA linear EST 07-NOV-2001
 DEFINITION 603624955F1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:5451828 5',
 mRNA sequence.
 ACCESSION BM045930
 VERSION BM045930.1 GI:16775197
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 779)
 NIH-MGC <http://mhc.nci.nih.gov/>,
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: DCTD/DRP
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM1942 row: m column: 13
 High quality sequence stop: 778.
 FEATURES
 source
 1. .779
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5451828"
 /clone_lib="NIH_MGC_40"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: prostate; Vector: pOTB7; Site:1: XhoI;
 Site:2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."
 BASE COUNT 201 a 170 c 143 g 265 t
 ORIGIN
 Query Match 23.3%; Score 532.2; DB 10; Length 779;
 Best Local Similarity 83.8%; Pred. No. 1.7e-98;
 Matches 614; Conservative 0; Mismatches 118; Indels 1; Gaps 1;
 QY 567 gggatgttctctgctgggcttctgttcgagggaacagcagaagcatccatcgtgttt 626
 Db 1 GCGATGCTCTCTCGTGGGCTCATGTTTGAGATAACGCGAAAGCTTCTATCATTTT 60

QY 627 ctcactcttcagcacagtgctgtcctgcagcagagaaacagccagcggtccaaag 686
 Db 61 GTTACTCTTTCAGTACATGTCTACCCCTCCAGCAACAGAAAGAACTACTGAATCAGTG 120
 QY 687 tcctctattattatgtgtcaaaagatttggccacggtttcttctacatgctgggtggca 746
 Db 121 TCCCTTTATTACTATGTCATCAAGATTTGGCTACTGTTTCTTCTTACATGCTAGTGGCG 180
 QY 747 atcattattcagccacaattcaggaaatgtgttggtataaaataaacagaagaatgcag 806
 Db 181 ATAATTTATTTCATGCCCTTAATTCAGAGATATATGTTGGATAAAATTAACAGCGCAATGCAC 240
 QY 807 ticaccaagcgaacaaacaaagtttaacagatctggctcagtttagtggtttctacttt 866
 Db 241 TTTCCAAAACAAACACAGCAAGTTTAAATGAATCTGGTCAGCTTAGTGGTGTCTACCTT 300
 QY 867 tttctgttatttggggcacattcttaactctctgaaaactgcctgtcagaccacaa 926
 Db 301 TTTGCCCTGTGTGGGGCACATTTCTCTCATCTCTGAAAACCTACATCTCAGACCCCACT 360
 QY 927 cttatatgaagcgctcgtccccaatagcatgacatttcaaatgaagtttttctacata 986
 Db 361 ATCTTATGGAGGGCTTATCCCCATAACCTGATGACATTTCAAAATGAAGTTTTTCTACATA 420
 QY 987 tccagttggcttactgtttcatgcttttctgaaacttcttccagaaacttctccagaaacccaaaa 1046
 Db 421 TCACAGCTGGCTTACTGGCTTTCATGCTTTTCTCGAACTTCTACTTCCAGAAACCAAAAA 480
 QY 1047 caagacatccctcgtcacaactgtctacattgttctacattggtcttccactcttccacattcgtgagct 1106
 Db 481 GAAGATATTCTCGTCAGCTTGTCTACATTTGGTCTTTACTCTTTCCACATTTGCTGGAGCT 540
 QY 1107 tatctctgtacttgaaatcatttgggactcttcttcttggactgcattatttctgtgaa 1166
 Db 541 TACCTTTTGAACCTGAATCATCTAGACCTGTTCTTCTGTGTGTACATTTATTTTCGTTGAA 600
 QY 1167 ttaactccacatgtgcgacctgttttacttagtgatgaaaaagaccagaa-aggcatt 1225
 Db 601 TTTCTTTTCCACATTTCCCGCTGTTTTATTATTAGCAATGAAAAGTATCAGAACAGAGATT 660
 QY 1226 atctctgtggccattgtgtttatctcttggtagactgtgacttgaattgttccgtact 1285
 Db 661 TTTCTCTGTGGCAGTTCTTTTAGTACTGGGAAGACTTCTGACTTTCATTTCTTTCAGTACT 720
 QY 1286 cactgttgggttt 1298
 Db 721 GACTGTGGGTTT 733
 RESULT 9
 LOCUS BE612552
 DEFINITION 601452065F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3855665 5',
 mRNA sequence.
 ACCESSION BE612552
 VERSION BE612552.1 GI:9894149
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 790)
 NIH-MGC <http://mhc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: DCTD/DRP
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be


```

Db 417 CCGTAATTCACAGATATAGTTGGATAAAATTAACAGGCGAATGCACITCTCCAAACAA 476
QY 820 aacaaacaagtttaacagagctgtgctagcttagtgctgtctactttttttctgtattt 879
Db 477 AACACAGCAAGTTAATGAATCTGCTAGCTTAGTGCCTTACCTTTTGGCTGTGTTT 536
QY 880 ggaggacatttatttaactctgaaactgcctgtcagacccaactcttatatgaagg 939
Db 537 GGGGCACATTCATCTCTGAAACTACATCTCAGACCCAACTATCTTATGAGGG 596
QY 940 ctgtcccatagcatgatgacatttcaaatgaagttttttctacatatccagttggctt 999
Db 597 CTTATCCCCATAACCTGATGACATTTCAATGAAGTTTTCCTACATATCACAGCTGGCTT 656
QY 1000 actgttttcatgcttttctcgaactctac-ttccagaaacaaacaaacaaacacacccct 1058
Db 657 ACTGGCTTCATGCTTTCTGTAACCTCTACTTTCCAGAAACCAAAAGAGATATTCCT 716
QY 1059 cgtcaactgtctacatggtcttcacctctccacattccacattcagagcttatctctgtac 1118
Db 717 CGTACGTTGCTACATGGTCTTTACCTCTCCACATTCCTGGAGCTTACCTTTTGAC 776
QY 1119 ttgaatcaatttggagcttctcttctgtgactgcatatttttggtaatttactttccac 1178
Db 777 TTGAATCATCTAGGACTTGTCTCTGGTGCTACATTAATTTTGTGGATTCTTTTCCGG 836
QY 1179 atgtgcgccttttttacttttag 1201
Db 837 CATTTCCCGCTGGTTAATTTTAG 859

RESULT 11
AUI31342
LOCUS AUI31342 NT2RP3 Homo sapiens cDNA clone NT2RP3002413 5', mRNA
DEFINITION
ACCESSION AUI31342
VERSION AUI31342.1 GI:10991696
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 905)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5', & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Location/Qualifiers
1..905
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP3002413"
/clone.lib="NT2RP3"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 214 a 221 c 212 g 255 t
ORIGIN 3 others
```

```

Query Match 22.2%; Score 508.2; DB 9; Length 905;
Best Local Similarity 81.8%; Pred. No. 1.4e-93;
Matches 597; Conservative 0; Mismatches 131; Indels 2; Gaps 1;

QY 465 ggaagcgccaccatgggctccgtaagaagagcaccacaagaccccccgcttctcagccag 524
Db 176 GCGGCCCCACCATGGCGATTCGCAAGAAAGACCAAGAGCCGCCAGCTGTGTGAGCCAC 235
QY 525 gaattcctctgcagaatcgcgcacatcgtctcctcgtcgtgggagtgcttctctcgtg 584
Db 236 GAATTCGCTCTCAGAAATCACGCGGACATCGTCTCTGTGTGGCGATGCTCTCTCTGCTG 295
QY 585 gggcttggttcgaagggaacagcagaatccatccatcgtgttttctcactcttcagcacagt 644
Db 296 GGGCTCATGTTTGAGATAACGGCAAAAGCTTCTATCATTTTGTACTCTTTCAGTACAAAT 355
QY 645 gtgtgctccctgcagcagaggaacacacgagggctcaaaagtcctctctattattatggt 704
Db 356 GTACCCCTCCAGCAACAGAGAAAGCAAGCTACTGAATCAGTGTCTCTTTTACTATGGC 415
QY 705 gtcaagatttggccacggttttctctacatcgtcgtggtggcaatcattatcattccaca 764
Db 416 ATCAAGAGATTGGCTACTGTTTTCTTCTACATGCTAGTGGCGATAAATATTTCATCCGTA 475
QY 765 attcaggaatatgttgataaaattacaagagaatgcaggtccacaaagcgaacaa 824
Db 476 APTCAAGAGTATATGTTGATAAAATTAAACAGGCGAATGCACCTTCTCCAAAACAAACAC 535
QY 825 aacaagttaacagagctcgtcagtttagtggttctacttttttctgtattttggggc 884
Db 536 AGCAAGTTTAATGAATCTGGTCAGCTTAGTGGCTTCTACCTTTTGGCTGTCTTTGGGC 595
QY 885 acattcatttaactctcgaagacgcctcgtcagacccaactcttatatggaagctcgt 944
Db 596 ACATTCAATTCATCTCTGAAACCTACATCTCAGACCCCAACATATCTTTTGGAGGGCTTAT 655
QY 945 ccccatagcatgacacatttcaaatgaagtttttctacatcccatccagttggtctactgg 1004
Db 656 CCCCATACCTGATGACATTTCAAATGAAGTTTCTCATATACAGAGCTGGCTTACTGG 715
QY 1005 ttctatgcttttccctgaactctactctccagaaacacaaacaaacagacatccctcgtcaa 1064
Db 716 CTTTCATGCTTTTCTGAACTCTACTTNCAGAAACCAAAAGAGATATTCCTCGTCAG 775
QY 1065 cttgtctacattggtcttccactctccacatctacatgagcttatctctgtacttgat 1124
Db 776 CTTGGCTACATTGGTCTTTACCTCTTTTACATTTGCTGGAGCTTACCTTTTGAACCTGAAT 835
QY 1125 catttgggactctctcttttggtaactgattattttgttgaaattactttccacacatg 1184
Db 836 CATCTAGGACTGGTCTCTGGG--CTACATTATTTGGTGAATTTCTTTTTCATATTNCC 893
QY 1185 ggcctggtttt 1194
Db 894 GCTGTTTATT 903

RESULT 12
BG547017
LOCUS BG547017 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4702154 5',
DEFINITION mRNA sequence.
ACCESSION BG547017
VERSION BG547017.1 GI:13545682
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 727)
NIH-MGC http://mgc.nci.nih.gov/.
```

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-i@mail.nih.gov
Tissue procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CMI538 row: 1 column: 03
High quality sequence stop: 714.

```

FEATURES
  Location/Qualifiers
    1..727
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:4702154"
      /clone_lib="NIH_MGC_77"
      /lab_host="DH10B (T1 phage-resistant)"
      /note="Organ: lung; Site: pDNR-LIB (Clontech); Site: 1:
      sfll1 (ggcgccctgcgcc); Site: 2: sfll1 (ggccattatgcc); 5' and
      3' adaptors were used in cloning as follows: 5' adaptor
      sequence: 5'-CAGCGCCATTATGGCC-3' and 3' adaptor sequence:
      5'-ATTCTGAGCCGAGCGCGCCGACATG-dt(30)BN-3' (where B = A,
      C, or G and N = A, C, G, or T). Average insert size 1.9
      kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
      by PCR. This library was enriched for full-length clones
      and was constructed by Clontech Laboratories (Palo Alto,
      CA). Note: this is a NIH_MGC Library."
    190 a 128 g 213 t
  BASE COUNT
    190 a 128 g 213 t

```

BASE COUNT 190 a 176 c 128 g 233 t

ORIGIN

Query Match 22.0%; Score 503.8; DB 10; Length 727;
Best Local Similarity 84.0%; Pred. No. 1.1e-92;
Matches 604; Conservative 0; Mismatches 112; Indels 3; Caps 3;

486	QY	cgtaagaagcagcaaacgaaccccccttctctcagcaggaaattcatctcgcagaaatcat	545
2	Db	CGCAAGAAAACGACCAAGAGCCCCCAGTGTGTGAGCCACGAAATTCGTCCTGCGAAGATCAC	61
546	QY	gcgagacatgctctcctgcgtggggatgtctctcctgcgtggggcttgcgttcgaggggaaca	605
62	Db	CGGACATCGTCTCCTGTGTGGCGATGGTCTTCTCTGGGCTCATGTTTGAGATACG	121
606	QY	gcgagaagcatccatcgtgttttctcaactcttcagcaacagtggtgcgtctccctgcagcagag	665
122	Db	GCAAAAGCTTCTATCATTTTGTGTACTTCTTCAGTACAAATGTACCCCTCCGACGACACAGAA	181
666	QY	gaacaagccacgggctcaaaagtcctctattattatgggtgcaaaagatttggccacaggtt	725
182	Db	GAAACAAGCTACTGAATCAGTGTCCCTTTATTACTATGGCATCAAAGATTTGGCTACTGTT	241
726	QY	ttcttctacatgctgggtggcaatcatttatctgcacaaattccaggaatatgtgttgtggaat	785
242	Db	TTCTTCTACATGCTAGTGGCGATAATTAATTCATGCCGTAATTCGAAGAGTATATGTTGGAT	301
786	QY	aaaattcaacagaaatgcagttcaccaacagcgaaaaacaaacaggtttaacgagctctggt	845
302	Db	AAAATTAAACGGCAATGCACCTCTCCAAAACAAAACACACGCAAGTTTAAATGAATCTGGT	361
846	QY	cagtttagtggttctacttttttcttgttatatttggggcacaattcattttaatctctgaa	905
362	Db	CAGCTTAGTCGGTCTACCTTTTGGCTGTGTTTGGGGCACATTCATTCTCATCTCTGAA	421
906	QY	aactgcctctcagaccacaactcttatgaaagctcgtcccatagacatgacattt	965
422	Db	AACTACATCTCAGACCACAACATACTATTATGAGGAGGCTTATCCCAATAACCTGATGACATTT	481
966	QY	caaatgaagtttttctacatatccacagttgggttactgggttttcatgtctttctcctgaact	1025
482	Db	CAAAATGAAGTTTTTCTACATATCACAGCTGGCTTACGTGGGTTTATGCTTTTCTCTGAACTC	541

Qy	1026	tacttcagaaaaacccccagagacatccctcgctcaactgtgtcta-cattggctcttca	1084
Db	542	TACTTCAGAAAAACAAAAAGAGATATCTCGTCAGCTTGCTACCAATTGGTCTTTA	601
Qy	1085	cctctccacattactggagcttacc-tcttgttacttgaatacattgggacctctctttt	1143
Db	602	CCTCTTCACAAATTCCTGGAGCTTACCTTTTGAACCTTGAATCATCTAGGACTGTGTTCTC	661
Qy	1144	tggctact-gcattattttgttgaattacttcccacatgtcggcctgttttactttag	1201
Db	662	TGGTCTAACAAATATTTCTTGGAAATTTCTTTCCACATTTCCCGCCTGTTTATCTAAG	720
RESULT	13		
LOCUS	BF969571	836 bp	linear
DEFINITION	602271922F1 NIH_MGC_84 Homo sapiens CDNA clone IMAGE:4360085 5',	EST	22-JAN-2001
ACCESSION	BF969571		
VERSION	BF969571.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs-r@mail.nih.gov		
	Tissue Procurement: ATCC		
	CDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	plate: LLAM10000 row: 1 column: 06		
	High quality sequence stop: 672.		

```

FEATURES
source
high quality sequence stop: 0/2.
Location/Qualifiers
1. .836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4360085"
/clone_lib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: adrenal gland; Vector: pCMV-Sport6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally; oligo-dt
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
197 a 211 c 176 g 252 t
BASE COUNT
ORIGIN

```

	Query Match	21.9%	Score 500.6;	DB 10;	Length 836;
	Best Local Similarity	80.7%;	Pred. No. 5e-92;		
	Matches 598;	Conservative 0;	Mismatches 139;	Indels 4;	Gaps 1;
Qy	465	ggaagccaccatggggctcgttaagaagagacccaagaacccccccggtcttcagccag	524		
Db	29	GGCGCCCCACCATGGCGGATTCCGAAGAAAAGACACCAAGAGCCCCCCCCAGTCTGAGCCAC	88		
Qy	525	gaattcatcttcgcagaatcatcgggacatcgtctcctcgtcgtggggatgtttcttcctcgtc	584		
Db	89	GAATTGCTCTGCAGAAATCACGGGACATCGTCTCCTGTGTGGCGATGGTCTTCTCTGCTG	148		
Qy	585	gggcttggtgttcgagggaacgacgaaagcatcccatcgtgtttctcaatctcttcagcacagt	644		
Db	149	GGGCTCATGTTTGAGTAACCGCAAAAGCTCTCATATTTTGTACTCTTCAGTACAAT	208		

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 17:09:47 ; Search time 74.4 Seconds
(without alignments)
550.890 Million cell updates/sec

Title: US-09-807-470-4

Perfect score: 1920

Sequence: 1 MGLRKSTKNPPVLSQEFIL.....VGVTNSNRVDCPPKRRKRS 369

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1920	100.0	369	21	AA198147 Human WAR-1 amino
2	1920	100.0	369	22	ABG12234 Novel human diagno
3	1920	100.0	369	22	AA870895 Human WAR-1 protei
4	1919	99.9	369	22	AA870895 Human polypeptide,
5	1463	76.2	363	21	AA198146 Rat WAR-1 amino ac
6	1463	76.2	363	22	AA870696 Rat WAR-1 protein
7	1392.5	72.5	416	21	AA843601 Human cancer assoc
8	1129.5	58.8	304	20	AA148434 Human prostate can
9	815	42.4	384	22	AA895708 Human reproductive
10	602.5	31.4	368	22	AB865847 Drosophila melanog
11	602.5	31.4	368	22	AB866114 Drosophila melanog

12	602.5	31.4	1575	22	AB867362 Drosophila melanog
13	493	25.7	125	21	AA800189 Human secreted pro
14	144	7.5	411	17	AA86810 Saccharomyces cere
15	143	7.4	394	20	AA100876 Human LAPH-1 prote
16	143	7.4	394	22	AA878909 Human protein seq
17	143	7.4	394	22	AA893884 Human protein sequ
18	138	7.2	191	21	AA824489 Human secreted pro
19	127	6.6	310	21	AA841779 Arabidopsis thalia
20	117.5	6.1	488	22	AA842028 Human polypeptide
21	117.5	6.1	536	22	AA879893 Human protein seq
22	116.5	6.1	280	21	AA844197 Arabidopsis thalia
23	116.5	6.1	296	21	AA844196 Arabidopsis thalia
24	116	6.0	380	20	AA100877 Human LAPH-2 prote
25	116	6.0	380	22	AA88560 Human hydrophobic
26	115	6.0	254	21	AA841780 Arabidopsis thalia
27	114.5	6.0	181	22	AA817201 Novel signal trans
28	114	5.9	266	21	AA843806 Human cancer assoc
29	113	5.9	158	22	AA817576 Novel signal trans
30	106.5	5.5	332	22	AA872076 Human olfactory re
31	105.5	5.5	484	22	AA84864 E. coli cellular p
32	105.5	5.5	816	20	AA827339 Group B Streptococ
33	104.5	5.4	150	21	AA834938 Human secreted pro
34	102	5.3	311	22	AA800782 Human apoptosis pr
35	102	5.3	692	22	AA835236 Human glycine tran
36	101	5.3	638	22	AA847065 Glycine transporte
37	101	5.3	692	22	AA835235 Human neurotransmi
38	100.5	5.2	339	22	AB860006 Drosophila melanog
39	100	5.2	414	22	AA823517 Human EST encoded
40	100	5.2	642	21	AA81943 Human GLYTLIKE pro
41	100	5.2	642	21	AA81944 Human GLYTLIKE pro
42	100	5.2	1898	22	AB825514 Novel human diagno
43	99.5	5.2	313	21	AA833404 Zea mays protein f
44	98.5	5.1	577	22	AB829227 Novel human diagno
45	97.5	5.1	262	22	AB866068 Drosophila melanog

ALIGNMENTS

RESULT 1
AA198147
ID AA198147 standard; Protein: 369 AA.
AC AA198147;
XX AC
XX 22-AUG-2000 (first entry)
XX DE Human WAR-1 amino acid sequence.
XX DE Endoplasmic reticulum; WAR-1; cancer cell proliferation inhibitor;
XX KW diagnosis; cancer; sarcoma; human.
XX OS Homo sapiens.
XX PN WO200022123-A1.
XX PD 20-APR-2000.
XX PF 13-OCT-1999; 99WO-JP05631.
XX PR 13-OCT-1998; 98JP-0290711.
XX PA (SUMU) SUMITOMO PHARM CO LTD.
XX PI Tohdoh N, Yoshima T, Komiya K, Tojo S, Nemoto K, Ishikawa H;
XX PR Okuyama H;
XX DR WPI: 2000-317980/27.
XX DR N-PSDB: AAA38013.
XX PT Endoplasmic reticulum protein WAR-1 which inhibits cancer cell proliferation for use in treatment and diagnosis of cancer including sarcomas of high malignancy -

XX Claim 1; Fig 2; 89pp; Japanese.
XX This sequence represents an endoplasmic reticulum protein (WAR-1) amino
CC acid sequence. The invention includes rat and human WAR-1 sequences,
CC expression vectors containing the DNA, cells transformed with the
CC expression vector, antibodies against WAR-1, and probes and primers which
CC hybridise to the DNA encoding WAR-1. The WAR-1 protein inhibits the
CC proliferation of cancer cells, and is used in the treatment and diagnosis
CC of cancers including highly malignant sarcomas.
XX
XX Sequence 369 AA;

Query Match 100.0%; Score 1920; DB 21; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.7e-214;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGLRSTKSTNPPVLSQEFILQNHADIVSCVGMFFLLGLVFEGETAEASIVFLTLQHSVAVP 60
Db 1 mglrskstknppvlsqefilqnhadivscvgmffllglvfegetaeasivfltlqhsvavp 60
QY 61 AAEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKQNFN 120
Db 61 aaeqatgskslYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKQNFN 120
QY 121 ESGQSVYFFSCIWGTFILISENCLSDPTLIWKARPHSMFTFQMKFFYISOLAYWPHAF 180
Db 121 esgqsvyffsciwgtfiliseNCLSDPTLIWKARPHSMFTFQMKFFYISOLAYWPHAF 180
QY 181 PELYFQKTKKODIPRLQVYIGLHFLHITGAYLLYLNHGLLLVHLYFVELLSHMCGLFY 240
Db 181 pelyfqtktkkODIPRLQVYIGLHFLHITGAYLLYLNHGLLLVHLYFVELLSHMCGLFY 240
QY 241 FSDEKYQGISLWAIIVFVLGRVLTIVSVLTVGFHLAGSQNRNPDALTGNNVLAAKIAV 300
Db 241 fsdekyqgisLWAIIVFVLGRVLTIVSVLTVGFHLAGSQNRNPDALTGNNVLAAKIAV 300
QY 301 LSSCTIQAYVTWNLITLWLQRWVEDSNIOASCMKKRSRKRRTENGVGVTENRVD 360
Db 301 lssctiqayvtwNLITLWLQRWVEDSNIOASCMKKRSRKRRTENGVGVTENRVD 360
QY 361 PPKRKEKSS 369
Db 361 ppkrkekss 369

RESULT 2
ABG12234
ID ABG12234 standard; Protein; 369 AA.
XX
XX AC ABG12234;
XX
XX DT 18-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #12225.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO2001/5067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX PA (HYSE-) HYSEQ INC.
XX

PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDB; AAS76421.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 42593; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 369 AA;

Query Match 100.0%; Score 1920; DB 22; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.7e-214;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGLRSTKSTNPPVLSQEFILQNHADIVSCVGMFFLLGLVFEGETAEASIVFLTLQHSVAVP 60
Db 1 mglrskstknppvlsqefilqnhadivscvgmffllglvfegetaeasivfltlqhsvavp 60
QY 61 AAEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKQNFN 120
Db 61 aaeqatgskslYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKQNFN 120
QY 121 ESGQSVYFFSCIWGTFILISENCLSDPTLIWKARPHSMFTFQMKFFYISOLAYWPHAF 180
Db 121 esgqsvyffsciwgtfiliseNCLSDPTLIWKARPHSMFTFQMKFFYISOLAYWPHAF 180
QY 181 PELYFQKTKKODIPRLQVYIGLHFLHITGAYLLYLNHGLLLVHLYFVELLSHMCGLFY 240
Db 181 pelyfqtktkkODIPRLQVYIGLHFLHITGAYLLYLNHGLLLVHLYFVELLSHMCGLFY 240
QY 241 FSDEKYQGISLWAIIVFVLGRVLTIVSVLTVGFHLAGSQNRNPDALTGNNVLAAKIAV 300
Db 241 fsdekyqgisLWAIIVFVLGRVLTIVSVLTVGFHLAGSQNRNPDALTGNNVLAAKIAV 300
QY 301 LSSCTIQAYVTWNLITLWLQRWVEDSNIOASCMKKRSRKRRTENGVGVTENRVD 360
Db 301 lssctiqayvtwNLITLWLQRWVEDSNIOASCMKKRSRKRRTENGVGVTENRVD 360
QY 361 PPKRKEKSS 369
Db 361 ppkrkekss 369

RESULT 3
AAB70695
ID AAB70695 standard; Protein; 369 AA.

XX AAB70695;
XX 17-MAY-2001 (first entry)
XX Human WAR-1 protein sequence SEQ ID NO:1.
XX
KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
KW endoplasmic reticulum membrane transportation; secretory protein;
KW cell membrane protein; cytosolic; CNS active; anti-allergic; cancer;
KW antirheumatic; nervous system disorder; immune disorder; allergy;
KW rheumatism; skeletal disorder.
XX
OS Homo sapiens.
XX
XX WO200114582-A1.
XX
XX 01-MAR-2001.
XX
XX 17-AUG-2000; 2000WO-JP05488.
XX
XX 20-AUG-1999; 99JP-0234764.
XX
XX (SUMU) SUMITOMO PHARM CO LTD.
XX
XX Tohdoh N, Okuyama H, Imamura M, Ishikawa H, Nemoto K;
XX
XX WPI: 2001-202940/20.
XX
XX N-PSDB; AAF74780, AAF74782.
XX
XX Transformation of a cell with separate vectors expressing the sense and
XX antisense strands of WAR-1 DNA for screening secretory and membrane
XX proteins expressed by the cell
XX
XX Claim 2; Page 58-60; 79pp; Japanese.
XX
XX The present invention describes a screening method for secretory and
XX membrane proteins consisting of transformation of a cell with separate
XX expression vectors for the sense and antisense RNA of DNA encoding an
XX endoplasmic reticulum membrane protein participating in endoplasmic
XX reticulum transport of proteins. Also described are: (1) secretory and
XX cell membrane proteins identified by the screening method; (2) drug
XX compositions containing these proteins; (3) host cells transformed by
XX the separate expression vectors of the method; and (4) the preparation
XX of secretory and cell membrane proteins by culture of the transformants.
XX The method can be used for the identification and preparation of
XX proteins for use in the treatment and prevention of diseases such as
XX cancer, disorders of the nervous system, immune disorders (including
XX allergies and rheumatism) and skeletal disorders. The present sequence
XX represents a specifically claimed human WAR-1 protein from the present
XX invention.
XX
XX Sequence 369 AA;
XX
Query Match 100.0%; Score 1920; DB 22; Length 369;
Best Local Similarity 100.0%; Pred No. 2.7e-214;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGLRKKSTKNPPVLSQEFILQNHADIVSCVGMFFLLGLVFEGTAESIVFLTLQHSVAVP 60
Db 1 mglrkkstknppvlsqefilqnhadivscvgmffllglvfgftaesaivfltlqhsavp 60
QY 61 AAEQATGSKSLYYGVKDLATVFFMLVAIIHATIQEYVLVDKINKRMQFTAKQNKFN 120
Db 61 aaeqatgskslYYGVKDLATVFFMLVAIIHATIQEYVLVDKINKRMQFTAKQNKFN 120
QY 121 ESGQFVFFFCIWGTFILISNCLSDPTLIWKARPHSNMTFQMKFFYISQLAYWFHAF 180
Db 121 esqgfvfffciswgtfllisencldsdptllwkarphsnmtfQMKFFYISQLAYWFHAF 180
QY 181 PELYFKTKQDIPROLVYIGLHLFHTIGAYLLYLNHLGLLLEVLVHYFVELLSHMGGLFY 240
Db 181 pelyfktkqdiprolvYIGLHLFHTIGAYLLYLNHLGLLLEVLVHYFVELLSHMGGLFY 240

Db 181 pelyfktkkdiprqlvYIGLHfhtigayllYlnhglilllvlyhfvellshmglfy 240
QY 241 FSDEKYOKGISLWAIIVFLIGRLVTLIVSVLTGVFHLAGSQNRNPDLTGNVNLAAKIAV 300
Db 241 fsdekykgislwaiivflgrlvtliivsvltvgfhlagsqnrnpdaltgnvnlvlaakiav 300
QY 301 LSSSCTTQAYVTWNLTILWLQRWVEDSNIAQCMKKRKRSSKKRRTENGVGVEVSNNRVDC 360
Db 301 lsssectiqayvtwnltliwlgrwvedsniaqcmkkrsskrtsrskkrtegvetsnrvdc 360
QY 361 PPKRKRKSS 369
Db 361 pprkrkekss 369
RESULT 4
AAM93265
ID AAM93265 standard; Protein; 369 AA.
XX
AC AAM93265;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polypeptide, SEQ ID NO: 2725.
XX
XX Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
XX Homo sapiens.
XX
XX EP1130094-A2.
XX
XX 05-SEP-2001.
XX
XX 07-JUL-2000; 2000EP-0114089.
XX
XX 08-JUL-1999; 99JP-0194486.
XX
XX 11-JAN-2000; 2000JP-0118774.
XX
XX 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI: 2001-524255/58.
XX N-PSDB; AAK94181.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -
XX
XX Claim 8; SEQ ID NO 2725; 1380pp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesizing the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is a polypeptide
XX encoded by a full length human cDNA of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.
XX
XX Sequence 369 AA;
XX
Query Match 99.9%; Score 1919; DB 22; Length 369;
Best Local Similarity 99.7%; Pred. No. 3.5e-214;
Matches 368; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGLRKKSTKNPPVLSQEFILQNHADIVSCVGMFFLLGLVFEGTAESIVFLTLQHSVAVP 60

```
|||||
Db 1 mglrkkstknppvlsqefilqnhadivscgmffllglvfegtaeasivfltlqhsavvp 60
QY 61 AAEQATGSKSLYYGVKDLATVFFVMLVAIIHATIQEYVLDKINKRMQFTKAKONKEN 120
Db 61 aaeqatgskslYYGVKDLATVFFVMLVAIIHATIQEYVLDKINKRMQFTKAKONKEN 120
QY 121 ESGQSFVFFSCIGTFTILISENCISDPTLIWKARPHSMMTFQMKFFYISOLAYWFHAF 180
Db 121 esqsfvffscigwtftilisencisdptliwkarpshsmmtfqmkffysqlaywfhaf 180
QY 181 PELYQKTKQDIPROLVYIGLHLFHTGAYLLYLNHLGLLLVLHYFVELLSHMCGLFY 240
Db 181 pelyqtktkqdiprqlvyiglhflhtgayllylnhlgllllvlhyfvelldshmcglfy 240
QY 241 FDEKQKQGISLWAIIVILGRVTLIVSVLTGCFHLAGSONRNPDALTCGNVNVLAAKIAV 300
Db 241 fdekqkqgislwaivilgrvltlivsvltvgfhlagsnrdaltgnvnlvlaakiav 300
QY 301 LSSSCTIQAYVTWNLTTLQRWVEDSNIOASCMKKRSRKKRTENGVGVSNSRVDC 360
Db 301 lssctiqayvtwnlttlqwrvedsnlqascmkkrrsrrsrrkrtengvgvetsnrvc 360
QY 361 PPKRREKSS 369
Db 361 pprkrekss 369

RESULT 5
AAY98146
ID AAY98146 standard; Protein; 363 AA.
AC AAY98146;
XX
DT 22-AUG-2000 (first entry)
XX
DE Rat WAR-1 amino acid sequence.
KW Endoplasmic reticulum; WAR-1; cancer cell proliferation inhibitor;
KW diagnosis; cancer; sarcoma; rat.
XX
OS Rattus norvegicus.
XX
PN WO200022123-A1.
XX
PD 20-APR-2000.
XX
PF 13-OCT-1999; 99WO-JP05631.
XX
PR 13-OCT-1998; 98JP-0290711.
XX
PA (SUMU ) SUMITOMO PHARM CO LTD.
XX
PI Tohdoh N, Yoshima T, Komiya K, Tojo S, Nemoto K, Ishikawa H;
PI Okuyama H;
XX
WPI; 2000-317980/27.
DR N-PSDB; AAA38012.
XX
PT Endoplasmic reticulum protein WAR-1 which inhibits cancer cell
PT proliferation for use in treatment and diagnosis of cancer including
PT sarcomas of high malignancy -
XX
XX Claim 1; Fig 2; 89pp; Japanese.
XX
CC This sequence represents an endoplasmic reticulum protein (WAR-1) amino
CC acid sequence. The invention includes rat and human WAR-1 sequences,
CC expression vectors containing the DNA, cells transformed with the
CC expression vector, antibodies against WAR-1, and probes and primers which
CC hybridize to the DNA encoding WAR-1. The WAR-1 protein inhibits the
CC proliferation of cancer cells, and is used in the treatment and diagnosis
CC of cancers including highly malignant sarcomas.
XX
```

```
SQ Sequence 363 AA;
Query Match 76.2%; Score 1463; DB 21; Length 363;
Best Local Similarity 75.8%; Pred. No. 3.7e-161;
Matches 279; Conservative 43; Mismatches 40; Indels 6; Gaps 5;

QY 1 MGLRKKSTKNPPVLSQEFILQNHADIVSCGMFFLLGLVFEGTAEASIVFLTLQHSVAVP 60
Db 1 mglrkkstknppvlsqefilqnhadivscgmffllglvfegtaeasivfltlqhsavvp 60
QY 61 AAEQATGSKSLYYGVKDLATVFFVMLVAIIHATIQEYVLDKINKRMQFTKAKONKEN 120
Db 61 aaeqatgskslYYGVKDLATVFFVMLVAIIHATIQEYVLDKINKRMQFTKAKONKEN 120
QY 121 ESGQSFVFFSCIGTFTILISENCISDPTLIWKARPHSMMTFQMKFFYISOLAYWFHAF 180
Db 121 esqsfvffscigwtftilisencisdptliwkarpshsmmtfqmkffysqlaywfhaf 180
QY 181 PELYQKTKQDIPROLVYIGLHLFHTGAYLLYLNHLGLLLVLHYFVELLSHMCGLFY 240
Db 181 pelyqtktkqdiprqlvyiglhflhtgayllylnhlgllllvlhyfvelldshmcglfy 240
QY 241 FDEKQKQGISLWAIIVILGRVTLIVSVLTGCFHLAGSONRNPDALTCGNVNVLAAKIAV 300
Db 241 fdekqkqgislwaivilgrvltlivsvltvgfhlagsnrdaltgnvnlvlaakiav 300
QY 301 LSSSCTIQAYVTWNLTTLQRWVEDSNIOASCMKKRSRKKRTENGVGVSNSRVDC 360
Db 301 lssctiqayvtwnlttlqwrvedsnlqascmkkrrsrrsrrkrtengvgvetsnrvc 360
QY 361 PPKRREKSS 369
Db 361 pprkrekss 369

RESULT 6
AAB70696
ID AAB70696 standard; Protein; 363 AA.
AC AAB70696;
XX
DT 17-MAY-2001 (first entry)
XX
DE Rat WAR-1 protein sequence SEQ ID NO:2.
XX
KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
KW endoplasmic reticulum membrane transportation; secretory protein;
KW cell membrane protein; cytosolic; CNS active; antiallergic; cancer;
KW antirheumatic; nervous system disorder; immune disorder; allergy;
KW rheumatism; skeletal disorder.
XX
OS Rattus sp.
XX
PN WO200114582-A1.
XX
PD 01-MAR-2001.
XX
PF 17-AUG-2000; 2000WO-JP05488.
XX
PR 20-AUG-1999; 99JP-0234764.
XX
PA (SUMU ) SUMITOMO PHARM CO LTD.
XX
PI Tohdoh N, Okuyama H, Imamura M, Ishikawa H, Nemoto K;
XX
WPI; 2001-202940/20.
DR N-PSDB; AAF74781.
XX
PT Transformation of a cell with separate vectors expressing the sense and
PT antisense strands of WAR-1 DNA for screening secretory and membrane
PT proteins expressed by the cell -
XX
```

PS Claim 3; Page 60-62; 79pp; Japanese.

XX The present invention describes a screening method for secretory and membrane proteins consisting of transformation of a cell with separate expression vectors for the sense and antisense RNA of DNA encoding an endoplasmic reticulum membrane protein participating in endoplasmic reticulum transport of proteins. Also described are: (1) secretory and cell membrane proteins identified by the screening method; (2) drug compositions containing these proteins; (3) host cells transformed by the separate expression vectors of the method; and (4) the preparation of secretory and cell membrane proteins by culture of the transformants.

CC The method can be used for the identification and preparation of proteins for use in the treatment and prevention of diseases such as cancer, disorders of the nervous system, immune disorders (including allergies and rheumatism) and skeletal disorders. The present sequence represents a specifically claimed rat WAR-1 protein from the present invention.

XX Sequence 363 AA;

SQ

Query Match 76.2%; Score 1463; DB 22; Length 363;
Best Local Similarity 75.8%; Pred. No. 3.7e-161;
Matches 279; Conservative 43; Mismatches 40; Indels 6; Gaps 5;

QY 1 MGLRRKTKNPPVLSOEFTLQNHADIVSCVGMFFLLGLVFEGETASIVFLTLQHSVAVP 60
DB 1 mglrkknarnppvlshemvqhdamvscvgmffvlgmfegtaemsivfltlqhgvvvp 60

QY 61 AAEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKONFEN 120
DB 61 -aeglpsgsrtylhygvkdlatvffymvvaaiihatiqeyvldklsrrlqltkgqknln 119

QY 121 ESGQFSVYFFSCINGTFTLISENCLSDPTLIWKARPHSMFTFQMKFFYISOLAYWFHAF 180
DB 120 eagqlsvrlyvsglwgmlasencldptllwksqphnmftfkmkffysqlaywfhsf 179

QY 181 PELYQKTKKODIPQLVYIGLHLEPHITGAYLLYNHLGLLLVLYFVELLSHMCGLFY 240
DB 180 pelyfqkrkqdpqgllygihlfhigayllynhlglllmhyavellsrvcsily 239

QY 241 PSDEYKQGISLWALVFLIGRLVTLIVSVLVGFHLAGSQNRNPALTCGNVNLAAKTAV 300
DB 240 fgedryqgslwplvifsgrlvtlsvvtvghlagt-nrngnalsgnvnvlaakiav 298

QY 301 LSSSCTIOAYVTWNLITLWLRWVEDSNIOASCMKKRSRKKETENGVGVTENRVDC 360
DB 299 lssscslqvtytwtltvwlqrledanlhv-cgrkrrsr-srkqten--gvnpenrids 354

QY 361 PPKRREKS 368
DB 355 ppkkka 362

RESULT 7
ID AAB43601
XX AAB43601 standard; Protein; 416 AA.
AC AAB43601;
DT 08-FEB-2001 (first entry)
DE Human cancer associated protein sequence SEQ ID NO:1046.
XX Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotrophic; antipsoriatic; angiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;

KW neurological disease; drug screening.

XX Homo sapiens.

OS

XX WO200055350-A1.

PN

XX 21-SEP-2000.

PD

XX 08-MAR-2000; 2000WO-US05882.

PF

XX 12-MAR-1999; 99US-0124270.

PR

XX (HUMA-) HUMAN GENOME SCI INC.

PA

PI Rosen CA, Ruben SM;

XX WPI; 2000-587533/55.

DR N-PSDB; AAC77810.

XX

PT Novel isolated nucleic acids comprising sequences encoding peptides useful for treating or diagnosing e.g. cancer -

PT

XX Claim 11; Page 1634-1636; 2352pp; English.

PS

XX AAC77607 to AAC78448 encode the human cancer associated proteins given in AAB43398 to AAB44239. The proteins can have activities based on the tissues and cells the genes are expressed in. Example of activities include: cytostatic; proliferative; vulnerable; immunomodulator; antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral; dermatological; neuroprotective; cardiac; thrombolytic; coagulant; nootropic; vasotrophic; antipsoriatic and angiogenic. The polynucleotides and polypeptides can be used for preventing, treating or ameliorating medical conditions and diagnosing pathological conditions. Polynucleotides, polypeptides, antibodies, agonists and antagonists from the present invention may be used to treat immune disorders by activating or inhibiting the proliferation, differentiation or mobilisation of immune cells, to treat disorders of haematopoietic cells, autoimmune disorders, allergic reactions, graft versus host disease and organ rejection, modulate haemostatic or thrombolytic activity, modulate inflammation, cancers, cardiovascular disorders, neurological disease and bacterial or viral infections. The peptides, nucleotides, antibodies, agonists and antagonists may be also used in drug screens. AAC78449 to AAC78457 and AAB44240 represent sequences used in the exemplification of the present invention.

XX

SQ Sequence 416 AA;

Query Match 72.5%; Score 1392.5; DB 21; Length 416;
Best Local Similarity 71.4%; Pred. No. 7e-153;
Matches 267; Conservative 44; Mismatches 58; Indels 5; Gaps 2;

QY 1 MGLRRKTKNPPVLSOEFTLQNHADIVSCVGMFFLLGLVFEGETASIVFLTLQHSVAVP 60
DB 43 mairkkstksppvlshemvqhdamvscvgmffvlgmfegtaemsivfltlqhvntlp 102

QY 61 AAEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKONFEN 120
DB 103 ateeqatesvlyyygiklatvffymvvaaiihatiqeyvldklnrrmfskthskfn 162

QY 121 ESGQFSVYFFSCINGTFTLISENCLSDPTLIWKARPHSMFTFQMKFFYISOLAYWFHAF 180
DB 163 esgqlsalfylfacwgtfllsenysdptllwrayphnmtfkmkffysqlaywlfhaf 222

QY 181 PELYQKTKKODIPQLVYIGLHLEPHITGAYLLYNHLGLLLVLYFVELLSHMCGLFY 240
DB 223 pelyfqkktkedipqlyvlylflhiagayllnlnhlgvllylhyfvefifhsrlyf 282

QY 241 PSDEYKQGISLWALVFLIGRLVTLIVSVLVGFHLAGSQNRNPALTCGNVNLAAKTAV 300
DB 283 fsnekyqgslwplvifsgrlvtlsvvtvghlagt-nrngnalsgnvnvlaakiav 342

Query Match	31.4%;	Score	602.5;	DB	22;	Length	368;
Best Local Similarity	36.9%;	Pred.	No. 4.4e-61;				
Matches	138;	Conservative	63;	Mismatches	132;	Indels	41;
Gaps							
QY	4	RKSTKPPVLSQEFILQNHADIVSCVMEFLGLGYPEGTAEASIVFLTLQHSVA-VPAA	62				
DB	9	rktsnkuppilishefvignhadliiscvamfvvgllmnestaafasafislhhnvsge	68				
QY	63	EEQATGSKSLYYGKDLAVFFVYMLVAIIIHATIQEYVLVDKINKRMQFTKAKQNFES	122				
DB	69	reqpygkptyiagikdycaiffytllccimhaiqefvidkiskklhlskflarfnes	128				
QY	123	GQFSVVFVFCINGCTFLIISENCLSDPFLTIWKARPHSMVTQMKFFYISOLATWFHAPPE	182				
DB	129	gqlvafyflsfgahvllkegylgvqaqwefpdpmpsfllhkfvfvqlaylhlmlpe	188				
QY	183	LYFO--KTRKQDIPROQVYI--GLHLFIHTGAYLLYNLHLLGLLVLHYFVELLSHMCGLF	239				
DB	189	lyfqiktkeeqpkvhsisgftl--ivlaytlsfqlalvltlthysellshvqli	246				
QY	240	--YPSDEKYOGKISLWVILGRVLTVLIVSVLTVGFHLGAGSONRNPDALTGNNVLA	297				
DB	247	gvfdreerlaklrvmnnavflirfatsvgitlyyvgivrs-----eqirakreak	290				
QY	298	IAVLSSTCTQAVYVWNLIITLWLQRWVEDSNIQASCMKKRKRSSKKRTENGVG	357				
DB	291	ilalgialqgylvifsfit-----eqirakreakreakialqtkkxp	336				
QY	358	VDCPP-----KRREKS	368				
DB	337	aktpkdkvkrkkes	350				
RESULT 12							
ID	ABB67362 standard; Protein; 1575 AA.						
XX	ABB67362;						
XX	26-MAR-2002 (first entry)						
XX	Drosophila melanogaster polypeptide SEQ ID NO 28878.						
XX	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.						
XX	Drosophila melanogaster.						
XX	WQ200171042-A2.						
XX	27-SEP-2001.						
XX	23-MAR-2001; 2001WO-US09231.						
XX	23-MAR-2000; 2000US-191637P.						
XX	11-JUL-2000; 2000US-0614150.						
XX	(PEKE) PE CORP NY.						
XX	Venter JC, Adams M, Li PWD, Myers EW;						
XX	WPI; 2001-656860/75.						
XX	N-PSDB; ABL11465.						
XX	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -						
XX	Disclosure; SEQ ID NO 28878; 21pp + Sequence Listing; English.						
XX	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention						
CC							

CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
CC	sequences (AB101840-ABL16175) and the encoded proteins
CC	(ABB5737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 1575 AA;
	Query Match 31.4%; Score 602.5; DB 22; Length 1575;
	Best Local Similarity 36.9%; Pred. No. 3.6e-60;
	Matches 138; Conservative 63; Mismatches 132; Indels 41; Gaps
QY	4 RKKSTKPNPVLISQEPILONHADIVSCVGMFLLGLVFEGETAEASIVFLTLQHSVA-VPAA 62
DB	1216 rktstknknpilisheviqnhadiiscvamfvvgimnestaafasafislhlnvsgepds 1275
QY	63 EQAFGSKSLYGYGVKDLATVFFYMLVAIIITHATIQEVLDKINKRMQFTKAKQKNFNES 122
DB	1276 reqpygkytyiagikdycaiffyrltcimhaiiqefvldkiskhlhskfklarfnes 1335
QY	132 GQFSVYFFSCIWGFTIILISENCLSDPTLIWKARPHSMWTQMFFYISQLAYWPHAPPE 182
DB	1336 gqlvafyilslfwgahvillkegylgqvaglvvegfdphgmsflhkfyfvvqlayylhmlpe 1395
QY	193 LYEQ--KTKQDIPQLVYI-GLHLFHTIGAYLLNLHGLLILLVHVFVLLSHMCGIF 239
DB	1396 lyfqlkikceeqgkxvhsisgfti--ivlaytisfqlaivllclhyfseilshvql 1453
QY	240 --YFSDERYQKGISLWAIYVFLGRVLTLIVSVLTGVFHLAGSQNRNPDALTCNVNVLAAK 297
DB	1454 gvfdceerlaklravnnavfflirfatvsgivgtlyyiggyrs----- 1497
QY	298 IAVLSSSTCIQAYVTWNLTITLWLRWVEDSNIQASCMKKRSRKSCKRTENGVGVEYSNR 357
DB	1498 llaigglialggylvfsfit-----eqllakreakreaklqlqtqkpp 1543
QY	358 VDCPP--KRREKS 368
DB	1544 aktpkdkvkrkkes 1557
	RESULT 13
	AA000189
ID	AA000189 standard; Protein; 125 AA.
XX	
AC	AA000189;
XX	
DT	06-OCT-2000 (first entry)
XX	
DE	Human secreted protein, SEQ ID NO: 4270.
XX	
KW	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation
KW	gene therapy; chromosome mapping.
XX	
OS	Homo sapiens.
XX	
PN	EP1033401-A2.
XX	
PD	06-SEP-2000.
XX	
PF	21-FEB-2000; 2000EP-0200610.
XX	
PR	26-FEB-1999; 99US-0122487.
XX	
PA	(GEST) GENSET.
XX	
PI	Dumas Milne Edwards J, Duclert A, Giordano J;
XX	

DR WPI; 2000-500381/45.
 XX N-PSDB; AAC00195.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 13; SEQ ID 4270; 71pp + CD-ROM; English.
 XX
 CC The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 XX
 SQ Sequence 125 AA;

Query Match 25.7%; Score 493; DB 21; Length 125;
 Best Local Similarity 76.6%; Pred. No. 5e-49;
 Matches 95; Conservative 15; Mismatches 14; Indels 0; Gaps 0;
 QY 1 MGLRKKSTKNPVLQSQFTLQNHADIVSCVGMFFLLGLVFEGETAEASIVFLTLQHSVAVP 60
 DB 1 mairkkstkppvlshfvlqphadivscvamvflglmfeltakasiifvtlqynvtip 60
 QY 61 AAEQATGSKSLYYGVKDLATVFFVFLVAIIHATIQEYVLDKINKRMQFTKAKQNKFN 120
 DB 61 ateeqatesvlyygiqlatvffvflvvaillhaviqeyimldkinrmxfsktxhskn 120
 QY 121 ESQ 124
 DB 121 esgq 124

RESULT 14
 AAR86810
 ID AAR86810 standard; Protein; 411 AA.
 XX
 AC AAR86810;
 XX
 DT 28-MAY-1997 (first entry)
 DE
 DE Saccharomyces cerevisiae LAG1 protein.
 XX
 KW Saccharomyces cerevisiae; LAG1; life-span limiting domain;
 KW life-span extending domain; stress tolerance; longevity;
 KW recombinant protein production.
 XX
 OS Saccharomyces cerevisiae.

Key Location/Qualifiers
 FT Domain 1..190
 FT /note= "life-span limiting domain"
 FT 194..411
 FT /note= "life-span extending domain"

PH WO9533834-A1.
 XX
 XX 14-DEC-1995.
 XX
 XX 02-JUN-1995; 95WO-US06725.
 XX
 XX 08-NOV-1994; 94US-0336031.
 PR 03-JUN-1994; 94US-0253875.

XX (RESE) RESEARCH CORP TECHNOLOGIES INC.
 XX Jazwinski SM;
 XX
 DR WPI; 1996-040238/04.
 DR N-PSDB; AAT07263.
 XX
 PT Eukaryotic LAG1 gene and protein - controls longevity, stress
 PT tolerance and reproductive capacity of eukaryotic cells, for
 PT improved prodn. of recombinant proteins.
 XX
 PS Claim 23; Page 100-102; 154pp; English.
 XX
 CC The Saccharomyces cerevisiae LAG1 gene product is composed of a
 CC life-span limiting domain (see AAR86812) and a life-span extending
 CC domain (see AAR86811). Overexpression of the LAG1 gene in older
 CC cells has a rejuvenating effect, which not only increases
 CC cellular life span, but also reproductive capacity and cellular
 CC tolerance to stress factors such as starvation and low pH. These
 CC cells may be used for the production of recombinant proteins. By
 CC increasing the life span of recombinant cells, the need for
 CC overexpression of recombinant gene products is avoided, and
 CC therefore any subsequent adverse effects on the host cell.
 XX
 SQ Sequence 411 AA;

Query Match 7.5%; Score 144; DB 17; Length 411;
 Best Local Similarity 22.6%; Pred. No. 1.1e-07;
 Matches 65; Conservative 49; Mismatches 112; Indels 62; Gaps 10;
 QY 33 FFLGLGVFEGETAEASIVFLTLQHSVAVP-----AAPEQATGSKSLYYGVKDLATVFFYM 87
 DB 88 ffilvcvy-----sayflsgnrtesnplhmfvaisyqvdgtds-yakgikdlisffvfm 140
 QY 88 LVAIIHATIQEYVLDKINKRMQFTKAKQNKFNESQFSVFYF-FSCINGTFLII-SENC 145
 DB 141 ifftfreflmdvprftvynvtsehrqkrmleqmaifycgvsqpgfglymyhshdlw 200
 QY 146 LSDPTLIWKARPHSMMTFQMKFFYISQLAYWFHAFPELYFOKTKODIPROLYVIGL-HL 204
 DB 201 lftkpmryttypvntpnplfkifvgaaifwaqgacvvlqllekprkdykelvfhiavl 260
 QY 205 FHITGAYLLYLNHLGLLLV-----LHYFVELLS-HMCGLFY----- 240
 DB 261 lliwssyvfhtkmglaivtmdvsgfflsksktlnynsvftpfvfglffvfiylrhv 320
 QY 241 -----FSDEKYOKGIGSLMAIVFILGRVLTII 266
 DB 321 vnirilwsvltfrehgnyvlnfatgqkwcwisl-pivfvlaalqlv 367

RESULT 15
 AAY00876
 ID AAY00876 standard; Protein; 394 AA.
 XX
 AC AAY00876;
 XX
 DT 21-MAY-1999 (first entry)
 XX
 DE Human LAPH-1 protein sequence.
 XX
 KW Longevity-assurance protein homologue; LAPH-1; human; cell proliferation;
 KW longevity-assurance protein homologue; LAPH-2; signal transduction;
 KW cell cycle regulation; apoptosis; cellular homeostatic pathway; aging;
 KW cancer; inflammation; autoimmune disease; infection;
 KW neurodegenerative disorder.
 XX
 OS Homo sapiens.
 XX
 XX WO9906558-A1.
 XX

Search completed: September 6, 2002, 17:09:48
Job time: 8277 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 17:10:38 ; Search time 28.82 Seconds
(without alignments)
312.736 Million cell updates/sec

Title: US-09-807-470-4

Perfect score: 1920

Sequence: 1 MGLRKKSTKNPVLSEFIL.....VGVERSNRVDPPKRKRSK 369

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_AA:*
- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/iaa/PCRUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	144	7.5	411	2	US-08-336-031-2
2	144	7.5	411	2	US-08-902-853-7
3	144	7.5	411	5	PCT-US95-06725-2
4	143	7.4	394	2	US-08-902-853-1
5	128	6.7	387	2	US-08-902-853-6
6	116	6.0	380	2	US-08-902-853-3
7	102	5.3	687	3	US-08-834-467-2
8	102	5.3	687	4	US-09-396-177-2
9	94.5	4.9	696	4	US-07-757-342D-4
10	93	4.8	638	1	US-08-295-814E-13
11	93	4.8	638	1	US-08-240-783B-2
12	93	4.8	638	4	US-09-084-813-2
13	93	4.8	638	4	US-09-343-361-13
14	93	4.8	638	5	PCT-US92-09662-2
15	86	4.5	193	2	US-08-336-031-6
16	86	4.5	193	5	PCT-US95-06725-6
17	86	4.5	2020	1	US-07-551-531-2
18	86	4.5	2485	5	PCT-US94-00198-1
19	86	4.5	2485	5	PCT-US94-00198-2
20	86	4.5	2818	1	US-08-510-284-1
21	86	4.5	2818	1	US-08-411-389-2
22	86	4.5	2818	2	US-08-449-933-2
23	86	4.5	2818	4	US-07-966-049A-2
24	86	4.5	2818	4	US-09-542-331-2
25	84	4.4	1956	4	US-08-843-417-2
26	83.5	4.3	362	2	US-08-902-853-5
27	83	4.3	317	2	US-08-790-572-3

28	83	4.3	317	2	US-09-213-398-3	Sequence 3, Appli
29	83	4.3	764	4	US-07-741-453A-60	Sequence 60, Appl
30	81	4.2	764	4	US-09-235-451-36	Sequence 36, Appl
31	80.5	4.2	363	3	US-09-046-086-2	Sequence 2, Appli
32	80	4.2	429	2	US-08-677-049-5	Sequence 5, Appli
33	80	4.2	1956	4	US-08-843-417-10	Sequence 10, Appl
34	78.5	4.1	317	2	US-08-619-362A-8	Sequence 8, Appli
35	77.5	4.0	494	1	US-08-689-974-3	Sequence 3, Appli
36	77.5	4.0	494	3	US-09-058-376-3	Sequence 3, Appli
37	77.5	4.0	1036	2	US-08-720-484A-5	Sequence 5, Appli
38	77.5	4.0	1036	4	US-08-953-823A-5	Sequence 5, Appli
39	77	4.0	395	3	US-08-781-891-73	Sequence 73, Appl
40	77	4.0	1269	3	US-08-781-891-74	Sequence 74, Appl
41	77	4.0	1432	3	US-08-781-891-71	Sequence 71, Appl
42	76.5	4.0	637	3	US-08-072-064-8	Sequence 8, Appli
43	76	4.0	617	1	US-07-879-617A-11	Sequence 11, Appl
44	76	4.0	617	1	US-08-301-722A-3	Sequence 3, Appli
45	76	4.0	617	1	US-08-240-783B-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-336-031-2
; Sequence 2, Application US/08336031
; Patent No. 5817782
; GENERAL INFORMATION:
; APPLICANT: Jazwinski, S. M.
; TITLE OF INVENTION: LAG1: A GENE FOR INCREASING THE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,031
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/253,875
; FILING DATE: 03-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 93032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-336-031-2

Query Match 7.5%; Score 144; DB 2; Length 411;
Best Local Similarity 22.6%; Pred. No. 3.7e-08;
Matches 65; Conservative 49; Mismatches 112; Indels 62; Gaps 10;
Qy 33 FFLGLVFEPTAEASIVFTLQHSVAVP-----AAEQATGSKSLYYGVKDLATVFFYM 87

Db 88 FFILVCVY-----SAYFLSGNRTESNPLHMFVAISVQVDGTD-SAYGKIKDLSFFFFYM 140
QY 88 LVAIIHATIQEYVLDKINKRMQFTRAKQKNESGQFSVFYF-FSCIMGTFFILI-SENC 145
Db 141 IFFTFLREFLMDVIRPFTVYLVNTSEHROKRMLEQMYAIFYCGVSPFGGLYIMYHSDLW 200
QY 146 LSDPTLLWKARPHSMMTFOMKFFYISOLAYWFAHFAFPPELYFQKTKKODIPROLVYIGL-HL 204
Db 201 LFKTKPWRYTPVITNPFLEKFIYLGQAFAWQAQACVLVLQLEKPRKDYKELVFHHIVTL 260
QY 205 FHITGAYLLNLHGLLLL-----LHYFVELLS-HMCGLEY----- 240
Db 261 LLWSSVVFHTKMGALAIYITMDVSDFLSLTKLNLVNSVFPFVGLFVFFWYILRHV 320
QY 241 -----FSDEKYQKGISLWAIWVILGRLVTLI 266
Db 321 VNIRILSVLTFRHEGNYVLNFAQYKWCISL-PIVFLVIAALQIV 367

RESULT 2
US-08-902-853-7
; Sequence 7, Application US/08902853
; Patent No. 5945330
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: HUMAN LONGEVITY-ASSURANCE PROTEIN HOMOLOGS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902,853
; FILING DATE: Herewith
; CLASSIFICATION: ?

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0345 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:

; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 541568
; US-08-902-853-7

Query Match 7.5%; Score 144; DB 2; Length 411;
Best Local Similarity 22.6%; Pred. No. 3.7e-08;
Matches 65; Conservative 49; Mismatches 112; Indels 62; Gaps 10;

QY 33 FFILGLVFECTAEASIVFLTLQHSVAVP-----AABEQATGSKSLYYGYVKDLATVFFYM 87
Db 88 FFILVCVY-----SAYFLSGNRTESNPLHMFVAISVQVDGTD-SAYGKIKDLSFFFFYM 140
QY 88 LVAIIHATIQEYVLDKINKRMQFTRAKQKNESGQFSVFYF-FSCIMGTFFILI-SENC 145
Db 141 IFFTFLREFLMDVIRPFTVYLVNTSEHROKRMLEQMYAIFYCGVSPFGGLYIMYHSDLW 200
QY 146 LSDPTLLWKARPHSMMTFOMKFFYISOLAYWFAHFAFPPELYFQKTKKODIPROLVYIGL-HL 204
Db 201 LFKTKPWRYTPVITNPFLEKFIYLGQAFAWQAQACVLVLQLEKPRKDYKELVFHHIVTL 260
QY 205 FHITGAYLLNLHGLLLL-----LHYFVELLS-HMCGLEY----- 240
Db 261 LLWSSVVFHTKMGALAIYITMDVSDFLSLTKLNLVNSVFPFVGLFVFFWYILRHV 320
QY 241 -----FSDEKYQKGISLWAIWVILGRLVTLI 266
Db 321 VNIRILSVLTFRHEGNYVLNFAQYKWCISL-PIVFLVIAALQIV 367

RESULT 3
PCT-US95-06725-2
; Sequence 2, Application PC/TUS9506725
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: LAG1: A GENE FOR INCREASING THE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06725
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/253,875 & 08/336,031
; FILING DATE: 03-JUN-1994 & 08-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 93032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-06725-2

Query Match 7.5%; Score 144; DB 5; Length 411;
Best Local Similarity 22.6%; Pred. No. 3.7e-08;
Matches 65; Conservative 49; Mismatches 112; Indels 62; Gaps 10;

QY 33 FFILGLVFECTAEASIVFLTLQHSVAVP-----AABEQATGSKSLYYGYVKDLATVFFYM 87
Db 88 FFILVCVY-----SAYFLSGNRTESNPLHMFVAISVQVDGTD-SAYGKIKDLSFFFFYM 140

Db 78 AICFACLLSPSLRPAEPIFLSYKQPDGS---YKGPKKADCFPIWIVTFAFRVIMD 134
QY 100 YVLDKI-----NKRQMTTKAKQKNFSGQSVFYFSCIW--GTFILISENCLSDPT 150
Db 135 YFRPVLWNGVNRKVI-----RFSQGG-YSPFYIL-CFWFLGLYIYRSNYSWNEE 186
QY 151 LWRKARPHSMFTQMKFFYISOLAYWFHAPPELYFQKTKK---QDIPROLVYVIGLHL--- 204
Db 187 KLFEDYPQYVMSPLFKAYVLIQGLFWLOQILVLHLEORRADHWQFAHHIVTCAILILSY 246
QY 205 ---FHTGAYLIYLNHGLLLL---VLHY--FVELLSHWCGLFYFSDKYQKGLSLWAI 255
Db 247 GFNFRVGNAILYIFDLSDYILSGGKMLYLGFGKICDYLFGIF-----VASWY 296
QY 256 V--FILGRVLTIVS---VLTGVGHLAGSONRNPDALTGNNVLAAKIAVLSSSCTIOAY 310
Db 297 SRHYLFESKILRVVVTNAPEIIGFHL-----DVPNGYIFNKPIYIAFILLFTLQL- 347
QY 311 VTWNLTILW 319
Db 348 ----LIYIW 352

RESULT 6
US-08-902-853-3
; Sequence 3, Application US/08902853
; Patent No. 5945330
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: HUMAN LONGEVITY-ASSURANCE PROTEIN HOMOLOGS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902,853
; FILING DATE: Herewith
; CLASSIFICATION: ?
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0345 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2NOT01
; CLONE: 493014
US-08-902-853-3

Query Match 6.0%; Score 116; DB 2; Length 380;
Best Local Similarity 20.2%; Pred. No. 5.7e-05;
Matches 56; Conservative 47; Mismatches 104; Indels 70; Gaps 11;
QY 107 KRMQTKAKQKNFSGQSVFYFSCIWGTFTILISENCLSDPTLIWKARP-HSMMTFQM 165
Db 121 RRNQRPSLLKKFKREASWRFTFYLIATAGMAVIVDKPWFYDMKKVWEGYPIQSTIPSOY 180
QY 166 KEFYISQLAYWFHAPPELYFQKTKQDIPROLVYICGLHFLHITGAYLL---VLNHL--G 219
Db 181 WYMYIELSFYWSLLFS--IASDVKRDKEQIIH-----HVATILISFSWFANVIRAG 232
QY 220 LLLLHLHYVEVLLSHMCGLFYFSDK--YQKISLWAIIVFILGRVLTIVSVL----- 270
Db 233 TLIMALHDSSDYLLSAAKMFNYAGWKNTCNNIFIVFAIVFIITRLVILPFWILHCTLVYP 292
QY 271 -----TVGFFHLAGSONRNPDALTGNNVLAAKIAVLSSSCTIOAYVTWNLTILWLO--- 321
Db 293 LELYPAFFGYFFNS-----MMGVQLLL-----HIFWAYLILRMAHKF 330
QY 322 ---RWVEDSNIQASCMMKKRSRKKRTENGVGIVETS 355
Db 331 ITGKLVED-----ERSDRETESSEGEAA 355

RESULT 7
US-08-834-467-2
; Sequence 2, Application US/08834467
; Patent No. 608015
; GENERAL INFORMATION:
; APPLICANT: Albert, Vivian
; TITLE OF INVENTION: Glycine Transporter
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows DEMONSTRATION Version 2.0D
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/834,467
; FILING DATE:
; CLASSIFICATION: 45
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 314572-109
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 687 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-834-467-2

Query Match 5.3%; Score 102; DB 3; Length 687;
Best Local Similarity 19.3%; Pred. No. 0.0058;
Matches 64; Conservative 42; Mismatches 91; Indels 134; Gaps 14;

QY 26 IVSCVGMFFLLGLVFEQTAEASIVFTLQHSVAVPAAEQATGSKSLYYGVKDLATVFF 85
Db 271 LLGCLGVSWL-----VFLCLIRGV-----KSSGKVVYF-----TATPPY 305
QY 86 YMLVAIIIIHATQEVYLDKINRM--QFTKAKONKFNESGQSFVYFFSCINGTFILIS- 142
Db 306 VVLTILFVRGVTFEGAFDGMVYLTPOWDKILEAKVWGDAAQIFVSLACAWGLITMAS 365
QY 143 -----ENCLSDPTLIWKARPHSMMTFQMKFFVYISOLAYWFHAFPELYFOKTKKQDIPROL 197
Db 366 YNKFHNCRDVSII-----SITNCA-----TS 388
QY 198 VYIGLHLFHITGAYLLYLNLHGLLILVLYHFVELLSHMCGLFYFSDKQKGIS----- 251
Db 389 VYAGVFIFSILG--FMANHLGVDV-----SRVADHGPGLAFVA--YPEALTLLPISP 436
QY 252 LWAIVF-----ILGRVTLIV-----SVLTGVPHLAGSONRNDP 285
Db 437 LWSLFFFMILLGLGTQFCLETLVTAIVDEVGNWILQKTYVTYTLGVAVAG----- 489
QY 286 ALTGNNVNLAAKIAVLSSSCTIQAVVTWNL 316
Db 490 -----FLGIPLTQAGIYWL 507

RESULT 8

US-09-396-177-2
; Sequence 2, Application US/09396177
; Patent No. 6251617
; GENERAL INFORMATION:
; APPLICANT: Albert, Vivian
; TITLE OF INVENTION: Glycine Transporter
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/396,177
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/834,467
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 314572-109
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 687 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-396-177-2

Query Match 5.3%; Score 102; DB 4; Length 687;
Best Local Similarity 19.3%; Pred. No. 0.0058;
Matches 64; Conservative 42; Mismatches 91; Indels 134; Gaps 14;

QY 26 IVSCVGMFFLLGLVFEQTAEASIVFTLQHSVAVPAAEQATGSKSLYYGVKDLATVFF 85
Db 271 LLGCLGVSWL-----VFLCLIRGV-----KSSGKVVYF-----TATPPY 305
QY 86 YMLVAIIIIHATQEVYLDKINRM--QFTKAKONKFNESGQSFVYFFSCINGTFILIS- 142
Db 306 VVLTILFVRGVTFEGAFDGMVYLTPOWDKILEAKVWGDAAQIFVSLACAWGLITMAS 365
QY 143 -----ENCLSDPTLIWKARPHSMMTFQMKFFVYISOLAYWFHAFPELYFOKTKKQDIPROL 197
Db 366 YNKFHNCRDVSII-----SITNCA-----TS 388
QY 198 VYIGLHLFHITGAYLLYLNLHGLLILVLYHFVELLSHMCGLFYFSDKQKGIS----- 251
Db 389 VYAGVFIFSILG--FMANHLGVDV-----SRVADHGPGLAFVA--YPEALTLLPISP 436
QY 252 LWAIVF-----ILGRVTLIV-----SVLTGVPHLAGSONRNDP 285
Db 437 LWSLFFFMILLGLGTQFCLETLVTAIVDEVGNWILQKTYVTYTLGVAVAG----- 489
QY 286 ALTGNNVNLAAKIAVLSSSCTIQAVVTWNL 316
Db 490 -----FLGIPLTQAGIYWL 507

RESULT 9

US-07-757-342D-4
; Sequence 4, Application US/07757342D
; Patent No. 6218509
; GENERAL INFORMATION:
; APPLICANT: IGARASHI, Masao
; MINEGISHI, Takashi
; NAKAMURA, Kazuto
; TITLE OF INVENTION: PROTEIN, DNA AND USE THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; CUSHMAN
; STREET: 130 Water Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,342D
; FILING DATE: 10-Sep-1991
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BUCKLEY, Linda M.
; REGISTRATION NUMBER: 31003
; REFERENCE/DOCKET NUMBER: 41226
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)523-3400
; TELEFAX: (617)523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 696 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-07-757-342D-4

Query Match 4.9%; Score 94.5; DB 4; Length 696;
Best Local Similarity 20.8%; Pred. No. 0.043;
Matches 75; Conservative 43; Mismatches 103; Indels 139; Gaps 17;

[illegible]

```

RESULT 10
US-08-295-814E-13
; Sequence 13, Application US/08295814E
; Patent No. 5658786
; GENERAL INFORMATION:
; APPLICANT: Smith, Kelli E.
; APPLICANT: Borden, Laurence A.
; APPLICANT: Hartig, Paul R.
; APPLICANT: Weinshank, Richard L.
; TITLE OF INVENTION: DNA ENCODING TAURINE AND GABA
; TITLE OF INVENTION: TRANSPORTERS AND USES THEREOF
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295,814E
; FILING DATE: DECEMBER 19, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40558-B-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-295-814E-13

```

Query Match	4.8%;	Score 93;	DB 1;	Length 638;
Best Local Similarity	18.9%;	Pred. No. 0.057;		

	Matches	63;	Conservative	42;	Mismatches	89;	Indels	140;	Gaps
Qy	26	IVSCVGMFFLLGLVFE	CTAEASIVFLTQHSV	AVPAEAEQATGSK	SLYYGVKDLATVFF	85			
		::: :::	::: :::	::: :::	::: :::				
Db	222	LLGCLGYSWV	-----	VVFLCLIRGV	-----	KSSGKVYVF	-----	TATFPY	256
Qy	86	YMLVAILIIHATIQEY	VLDKINKRMQFTKAK	QNKFNES	-----	CQFSVFFVFFSC	IGCTGFFIL	140	
		::: :::	::: :::	::: :::	::: :::				
Db	257	VVLTLFVRGVTLEGA	FTGI	---MYLTLPKWDK	ILIAKVMGDAASQ	IFYSLGCAWGGLIT	313		
Qy	141	IS	-----	ENCLSDPTLI	IKARPHSMMT	FQMKFFYISQLAY	WHPAFPELYFQKTKQDIP	194	
		::: :::	::: :::	::: :::	::: :::				
Db	314	MASYNKPHN	CYRDSVII	-----	SITNCA	-----		337	
Qy	195	RQLVYIGLHLPHTIG	AVLLYNHLGLLLV	LVHVFVELLSHM	CGLFVPSDEKVKOGIS	251			
		::: :::	::: :::	::: :::	::: :::				
Db	338	-TSYVAGFVIFSI	LG	---FMANHLGVDV	-----	SRVADHGPGLAEVA	---YPEALTLLP	384	
Qy	252	--LWAEV	VF	-----	ILGLRLVTLV	-----	SVLTATWGPHLAGSQNR	282	
		::: :::	::: :::	::: :::	::: :::				
Db	385	ISPWLSLFFFM	LILGLGTFQCL	ETFLVTAIVDE	GVNEWILOKKTYVT	ILGVAVAG	----	440	
Qy	283	NPDALTGNVN	VLAAKIAVLSS	SCTCIQAYVTWNLI	316				
		::: :::	::: :::	::: :::	::: :::				
Db	441	-----	FLGLIPTLSQAGI	WMLLL	458				

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match
4.8%; Score 93; DB 1; Length 638;

QY 283 NPDALTGNVNVLAAKIAVLSSSCTIQAYVYWNLI 316
Db 441 -----FLIGIPLTSQAGIYWLIL 458

RESULT 14
PCT-US92-09662-2
; Sequence 2, Application PC/TUS9209662
; GENERAL INFORMATION:
; APPLICANT: Smith, Kelli
; APPLICANT: Borden, Laurence A.
; APPLICANT: Brantchek, Theresa
; APPLICANT: Hartig, Paul R.
; APPLICANT: Weinsbank, Richard L.
; TITLE OF INVENTION: DNA ENCODING A GLYCINE TRANSPORTER AND USES THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/09662
; FILING DATE: 19921112
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 1795/39875-A-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 977-9809
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-09662-2

Query Match 4.8%; Score 93; DB 5; Length 638;
Best Local Similarity 18.9%; Pred. No. 0.057;
Matches 63; Conservative 42; Mismatches 89; Indels 140; Gaps 15;
QY 26 IVSCVGNFFLLGLVFGCTAASIVFLTLQHSVAVPAEAEQATGSKSLYYGVKDLATVFF 85
Db 222 LLGLGVSWY-----VVFCLIRGV-----KSSGKVYF-----TATFPY 256
QY 86 YMLVAIIHATIQEYVLDKINKRMQFTKAKONKNES-----GQFSVYFFSCIGCTFIL 140
Db 257 VVLTILFVRGVTLGAPTGI---MYLTPKWDKILCAKVWGDAAQIFSLGCANGGLIT 313
QY 141 IS-----ENCLSDPTLWINKARPHSMFTOMKFYISQLAYWHPAPPFLYFQKTKQDIP 194
Db 314 MASYNKFNHCYRDSVII-----SITNCA----- 337
QY 195 RQLVYGLHLFHTGAVLLYLNHLGALLLVHYFVVELLSHMCGLFYFSDEKYOKGIS--- 251
Db 338 -TSYAGFVIFSILG---FMANHGLVDV-----SRVADHGPGLAEVA---YPEALTLLP 384
QY 252 ---LWAIIVF-----ILGRVTLIV-----SVLTWGFHLGASQNR 282
Db 385 ISPLNSLLFFMLILLGLGTQFCLETLVTAIVDEGVNEWILQKTYVTILGVAVAG----- 440

QY 283 NPDALTGNVNVLAAKIAVLSSSCTIQAYVYWNLI 316
Db 441 -----FLIGIPLTSQAGIYWLIL 458

RESULT 15
US-08-336-031-6
; Sequence 6, Application US/08336031
; Patent No. 5817782
; GENERAL INFORMATION:
; APPLICANT: Jazwinski, S. M.
; TITLE OF INVENTION: LAG1: A GENE FOR INCREASING THE
; TITLE OF INVENTION: LONGEVITY OF EUKARYOTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,031
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/253,875
; FILING DATE: 03-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 93032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 193 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-336-031-6

Query Match 4.5%; Score 86; DB 2; Length 193;
Best Local Similarity 26.5%; Pred. No. 0.062;
Matches 27; Conservative 18; Mismatches 45; Indels 12; Gaps 3;
QY 33 FFLGLVFGCTAASIVFLTLQHSVAVP-----AAEQATGSKSLYYGVKDLATVFFYM 87
Db 88 FFILGVY-----SAYFLSGNRTESNPLHMFVAISYQVDGTDG--YAKGIKDLFSFFYFM 140
QY 88 LVAAIIHATIQEYVLDKINKRMQFTKAKONKNESGQFSVFX 129
Db 141 IFFTFLREFLMDVIRPFTVYLVNVTSEHQRKRMLEQMIAYFI 182
Search completed: September 6, 2002, 17:10:39
Job time: 7973 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 17:11:51 ; Search time 52.02 Seconds
(without alignments)
681.603 Million cell updates/sec

Title: US-09-807-470-4

Perfect score: 1920

Sequence: 1 MGLRKKSTKNPPVLSQEFIL.....VGVETSNRVDPPKREKSS 369

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1405.5	73.2	374	2 S21736	translocating chai
2	1392.5	72.5	374	2 S30034	translocating chai
3	494	25.7	371	2 T19417	hypothetical prote
4	487	25.4	373	2 T19419	hypothetical prote
5	144	7.5	411	2 S46800	LAG1 protein - yea
6	128	6.7	390	2 T38012	longevity-assuranc
7	123.5	6.4	418	2 S30134	hypothetical prote
8	116.5	6.1	357	2 T40389	longevity assuranc
9	115	6.0	614	2 B84949	NADH dehydrogenase
10	111.5	5.8	360	2 T27324	hypothetical prote
11	108.5	5.7	476	2 AG1051	probable transport
12	106.5	5.5	331	2 T20916	hypothetical prote
13	106.5	5.5	459	2 T11411	NADH dehydrogenase
14	106.5	5.5	484	2 A86116	hypothetical prote
15	105.5	5.5	484	2 D65230	hypothetical prote
16	105.5	5.5	484	2 A98275	hypothetical prote
17	102	5.3	638	2 T77912	glycine transporte
18	102	5.3	692	2 I57956	glycine transporte
19	101	5.3	347	2 A53306	dipeptide transpor
20	100.5	5.2	614	2 A59845	Na+/H+ antiporter
21	99.5	5.2	507	2 F37765	hypothetical prote
22	99	5.2	740	2 AB1260	probable integral
23	99	5.2	802	2 S64153	probable membrane
24	98.5	5.1	308	2 H86268	hypothetical prote
25	98.5	5.1	391	2 S07743	ubiquinol--cytochr
26	98.5	5.1	517	2 S21042	cytochrome-c oxida
27	98	5.1	387	2 S07749	hypothetical prote
28	97.5	5.1	506	1 D64048	iron (III) ABC tra
29	96.5	5.0	469	2 D70048	ABC transporter (a

30 96 5.0 1123 2 T48586
31 95.5 5.0 534 2 B96642
32 95.5 5.0 835 2 T15177
33 95.5 5.0 1564 2 T27121
34 95 4.9 499 2 H70453
35 94.5 4.9 489 2 C65147
36 94.5 4.9 489 2 H91174
37 94.5 4.9 489 2 H86020
38 94.5 4.9 696 2 A41344
39 94 4.9 214 2 D83913
40 94 4.9 512 2 G90399
41 93.5 4.9 394 2 T38922
42 93.5 4.9 493 2 S78183
43 93.5 4.9 648 2 B84139
44 93 4.8 633 2 I58140
45 93 4.8 638 2 JH0673

RESULT 1
S21736
translocating chain-associating membrane protein - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 24-Sep-1999
C:Accession: S21736
R:Goerlich, D.; Hartmann, E.; Prehn, S.; Rapoport, T.A.
Nature 357, 47-52, 1992
A:Title: A protein of the endoplasmic reticulum involved early in polypeptide transla
A:Reference number: S21736; MUID:92244357
A:Accession: S21736
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-374 <GOE>
A:Cross-references: EMBL:X63678; NID:g941; PIDN:CAA45217.1; PID:g942
C:Superfamily: translocating chain-associating membrane protein
C:Keywords: transmembrane protein

Query Match 73.2%; Score 1405.5; DB 2; Length 374;
Best Local Similarity 72.5%; Pred. No. 1.6e-115;
Matches 271; Conservative 40; Mismatches 58; Indels 5; Gaps 2;

Qy 1 MGLRKKSTKNPPVLSQEFILQNHADIVSCVGMFFLLGLVFECTAEASIVFLQLQHSVAVP 60
Db 1 MAIRKSTKSPVLSHEFILQNHADIVSCVAMVLLGLMFEITAKASIIFFVTLQYNVTL 60
Qy 61 AAEEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKONFN 120
Db 61 ATEEQATTESTSLYYGKDLATVFFYMLVAIIHATIQEYVLDKINKRMHFSKTKHNF 120
Qy 121 ESGQSVFFYFSCINGTIFILISENCLSDPTLIWKARPHSMFTQMKFFYSOLAYWFHAF 180
Db 121 ESGQSAFVFFSCINGTIFILISENCLSDPTLIWKARPHSMFTQMKFFYSOLAYWFHAF 180
Qy 181 PELYFQTKKODIPQLVYIGLPHFTGAYLLYNHLGLLLVLLVHYFVEFLSRFLY 240
Db 181 PELYFQTKKEDIPQLVYIGLPHFTGAYLLYNHLGLLLVLLVHYFVEFLSRFLY 240
Qy 241 FSDEKYQKISLWAVFIIIGRLVTLIVSVLTIVGFHLAGSQNRNPDALTCNVNLAIAV 300
Db 241 FSDEKYQKISLWAVFIIIGRLVTLIVSVLTIVGFHLAGSQNRNPDALTCNVNLAIAV 300
Qy 301 LSSSTCIQAVYVNLITLWLQWEDSNIQASCMKKK----RSRSKKRTENG-VGVETS 355
Db 301 LASICITQAFMMKPIFNQLRWREHSTFQADPVVKKKPTVTGKRSRKGTEGVNGTVTS 360
Qy 356 NRVDCTPPRKREKSS 369
Db 361 NGADSPRRNRKESKSS 374

```
RESULT 2
S30034
translocating chain-associating membrane protein - human
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-Sep-1999
C:Accession: S30034
R:Goerlich, D.; Hartmann, E.; Prehn, S.; Rapoport, T.A.
Nature 357, 47-52, 1992
A:Title: A protein of the endoplasmic reticulum involved early in polypeptide translocation
A:Reference number: S21736; MUID:92244357
A:Accession: S30034
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-374 <OE>
A:Cross-references: EMBL:X63679; NID:g37264; PIDN:CAA45218.1; PID:g37265
C:Superfamily: translocating chain-associating membrane protein

Query Match 72.5%; Score 1392.5; DB 2; Length 374;
Best Local Similarity 71.4%; Pred. No. 2.2e-114;
Matches 267; Conservative 44; Mismatches 58; Indels 5; Gaps 2;

QY 1 MGLRKKSTKPPVLSQEFILQNHADIYSCVGMFFLLGLVFEGETAEASIVFLTLQHSVAVP 60
DB 1 MAIRKSTKPPVLSHBEFVLQNHADIYSCVAMVLLGLMFEITAKASIIIFVTLQYNVTL 60
QY 61 AAEQATGSKSLYYGYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKON 120
DB 61 ATEQATGESVLSYYGKDLATVFFYMLVAIIHATIQEYVLDKINKRMHFSKTHSKFN 120
QY 121 ESGQFSVFFSCITWGTFFILISENCLSDPTLIWKARPHSMFTQMKFFYISQLAYWPHAF 180
DB 121 ESGQLSAFYLFACVWGTFILISENYSIDPTILWRAYPHNLMTQMKFFYISQLAYWPHAF 180
QY 181 PELFYQTKKQDIPROLVYIGLHLFHTGTAYLGLLNLHGLLLVHLHVFVLLSHMCGFY 240
DB 181 PELFYQTKKEDIPROLVYIGLYLHAGAYLLNLHGLVLLVHLHVFVFLHISLFLY 240
QY 241 FSDKYQKGLSLAIVFILGRVLTIVSVLTGPHLAGSNRNPDLTGNVNVLAARIAV 300
DB 241 FSNKYQKGLSLAVLFLVGLRLTLISVLTGPHLAGRAENQKLDFTGPNVLAARIAV 300
QY 301 LSSSCTQAYVTWNLITLWLRQWVEDSNIOQSCMKKK-----RSRKKRTENG-VGYETS 355
DB 301 LASICVTQAFMMWKFINFQLRRWRHSAFOAPAVKKKPTVTTKGRSSKKGTEGVNGTLTS 360
QY 356 NRVDPCPKRKKSS 369
DB 361 NVADSPRKKKKSS 374

RESULT 3
T19417
hypothetical protein C24F3.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T19417
R:McMurray, A.
submitted to the EMBL Data Library, April 1998
A:Reference number: Z19122
A:Accession: T19417
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-371 <WTL>
A:Cross-references: EMBL:AL022716; PIDN:CAA18770.1; GSPDB:GN00022; CESP:C24F3.1a
A:Experimental source: clone C24F3
C:Genetics:
A:Gene: CESP:C24F3.1a
A:Map position: 4
A:Introns: 114/3; 158/2; 363/3
C:Superfamily: translocating chain-associating membrane protein
```

```
Query Match 25.7%; Score 494; DB 2; Length 371;
Best Local Similarity 32.4%; Pred. No. 1.3e-35;
Matches 119; Conservative 74; Mismatches 144; Indels 30; Gaps 9;

QY 2 GLRKKSTKPNPVLSSQEFILQNHADIYSCVGMFFLLGLVFEGETAEASIVFLTLQ-----HSV 57
DB 7 GSKASKKPPQPPILSHSEFIQNHGDMSCVVMVIVGLMFFLTHLSLSLFIAPQYNGTYTV 66
QY 58 AVPAEEQATGSKSLYYGYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKON 117
DB 67 AVEQGQEREVHG-----YLSGILDLPALFFYSCVWIVHVVAVQVEYGLDKISKTHLSKVSTF 123
QY 118 KFNESGQFSVFFSCITWGTFFILISENCLSDPTLIWKARP--HSMWTFOMKFFYISQL 173
DB 124 KFGESQHOMFEFTVYS-IAHAFYIVSERLEDFSEVKVSWLGYPTERHVMSAAAYKLYFIQI 182
QY 174 AYWFHAFPELYFOKTKKQDIPROLVYIGLHLFHTGTAYLGLLNLHGLLLVHLHVFVELLS 233
DB 183 SYWIHQPEFYLOKLRDEIRQKSVQAILHAFISIAFFNFTRVGLALITLEYITOLIF 242
QY 234 HMCGLFYFSDKQKGIS-----LWAIIVFILGRVLTIVSVLTGPHLAGSNRNPDLT 288
DB 243 HIARFAHVG---RKGLSDPAFLNGSFVLRVLSIIIAVMTWYGLRQAESPFVDISA 299
QY 289 GNVNVLAAKIAVLSSSCTQAYVTWNLITLWLRQWVEDSNIOQSCMKKKRS-----RS 341
DB 300 GNFTAVIRLNVLLAVLLQLFLYSPVFMHGRFRESN---AKKEKKKSAAAAAAPPKK 356
QY 342 SKRKTEN 348
DB 357 EKKRQDS 363

RESULT 4
T19419
hypothetical protein C24F3.1b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T19419
R:McMurray, A.
submitted to the EMBL Data Library, April 1998
A:Reference number: Z19122
A:Accession: T19419
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-373 <WTL>
A:Cross-references: EMBL:AL022716; PIDN:CAA18772.1; GSPDB:GN00022; CESP:C24F3.1b
A:Experimental source: clone C24F3
C:Genetics:
A:Gene: CESP:C24F3.1b
A:Map position: 4
A:Introns: 114/3; 158/2; 365/3
C:Superfamily: translocating chain-associating membrane protein

Query Match 25.4%; Score 487; DB 2; Length 373;
Best Local Similarity 32.4%; Pred. No. 5.2e-35;
Matches 120; Conservative 73; Mismatches 143; Indels 34; Gaps 10;

QY 2 GLRKKSTKPNPVLSSQEFILQNHADIYSCVGMFFLLGLVFEGETAEASIVFLTLQ-----HSV 57
DB 7 GSKASKKPPQPPILSHSEFIQNHGDMSCVVMVIVGLMFFLTHLSLSLFIAPQYNGTYTV 66
QY 58 AVPAEEQATGSKSLYYGYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKON 117
DB 67 AVEQGQEREVHG-----YLSGILDLPALFFYSCVWIVHVVAVQVEYGLDKISKTHLSKVSTF 123
QY 118 KFNESGQFSVFFSCITWGTFFILISENCLSD-----PTLIWKARP--HSMWTFOMKFFYI 170
DB 124 KFGESQHOMFEFTVYS-IAHAFYIVSER-LEDFSEVKVSWLGYPTERHVMSAAAYKLYFI 181
QY 171 SOLAYWFHAFPELYFOKTKKQDIPROLVYIGLHLFHTGTAYLGLLNLHGLLLVHLHVFVE 230
```


Db 182 FOISYWIHOFPEFYLOKLRDEIROKSVQAILHIAFISIAFYFNFTRVGLALTILEYITQ 241
QY 231 LLSHMCGLFYFSDKXKGIS-----LWAIVFLLGLRLTLIVSVLTGPHLAGSQNRNDP 285
Db 242 LIFHARFAHVG---RKLSDPAKFLNGSVFLVRLGSIITAVMTFTWGLRQAESPFVD 298
QY 286 ALTGNNVLAIAKIAVLSSTCTQAYVTNLTILWLQWVEDSNIQASCMKKRS-----339
Db 299 ISAGNFNTAVIRNLVLLAVVLQLFLLYSVFVHMGFRFRESN---AKKEKKSAANAARV 355
QY 340 -RSSKKRTEN 348
Db 356 PKRKKRQDS 365

RESULT 5
S46800
LAG1 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YHL003c
C:Species: Saccharomyces cerevisiae
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 23-Mar-2001
R:Favetto, T.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of S. cerevisiae cosmid 9780.
A:Reference number: S46797
A:Accession: S46800
A:Molecule type: DNA
A:Residues: 1-411 <FAV>
A:Cross-references: EMBL:U10555; NID:4500813; PIDN:AAB68429.1; PID:4500820; MIPS:YHL003c
R:D'Amelio, N.P.; Childress, A.M.; Franklin, D.S.; Kale, S.P.; Pinswasdi, C.; Jazwinski, J. Biol. Chem. 269, 15451-15459, 1994
A:Title: Cloning and characterization of LAG1, a longevity-assurance gene in yeast.
A:Reference number: A54012; MUID:94253121
A:Accession: A54012
A:Molecule type: DNA
A:Residues: 1-172, 'IV', 175-219, 'C', 221-300, 'TEISGI', 314, 'EKQE', 315, 'DSNDNPTE', 324, 'A', 324
'A', 381, 'AGOR', 386, 'L', 388, 'NRLARNEK' <DAM>
C:Genetics:
A:Gene: SGD:LAG1
A:Cross-references: SGD:S0000995; MIPS:YHL003c
A:Map position: 8L
C:Function:
A:Description: Involved in germination of longevity
C:Superfamily: hypothetical protein YKL008c
C:Keywords: transmembrane protein

Query Match 7.5%; Score 144; DB 2; Length 411;
Best Local Similarity 22.6%; Pred. No. 6.8e-05;
Matches 65; Conservative 49; Mismatches 112; Indels 52; Gaps 10;

QY 33 FFLGLVFGTAESAIVFTLQHSVAVP-----AAEEQATGSKSLYYGVKDLATVFFYM 87
Db 88 FFILVCVY-----SAYFLSGNRTESNPLHMFVAISYQVGDGDS-YAKGKDLSPFFYM 140
QY 88 LVAILIHATIQEYVLKINKRQFTKAKONKFNESQFSVFFY-FSCWTGTFILI-SENC 145
Db 141 IFTFLREFLMDVIRPFVYLVNVTSEHQRKRMLEQMAIFYCGVSGPGLYIMYHSDIW 200
QY 146 LSDPLIWKARPHSMMTFOMKFFYSOLAYNFAHPELYFQKTKQDIPRLQVYIGL-HL 204
Db 201 LFKTPMYTYPIYNPFLEKIFYLQAAFWAQACVQLVLEKPRKDYKFLVHHIVTL 260
QY 205 FHTGAYLLYLNHLGLLLV-----LHYFVELLS-HMCGIFY-----240
Db 261 LLIWSSYVFHETKMGLAIYITMDVSDFFLSLSTLNLYNSVTFPFVGLFVFFWYLRHV 320
QY 241 -----FSDKDYKIGSLMAIFILGRVLTIL 266
Db 321 VNIRILWSVLTFRHGNVNLNFAATQYKWCISL-PIVFVLLAALQLV 367

RESULT 6
T38012
longevity-assurance protein 1 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38012
R:Churcher, C.M.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z21761
A:Accession: T38012
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-390 <CHU>
A:Cross-references: EMBL:Z99258; PIDN:CAB16359.1; GSPDB:GN00066; SPDB:SPAC1A6.09c
A:Experimental source: strain 972h; cosmid clA6
C:Genetics:
A:Gene: lag1; SPDB:SPAC1A6.09c
A:Map position: 1

Query Match 6.7%; Score 128; DB 2; Length 390;
Best Local Similarity 22.0%; Pred. No. 0.0016;
Matches 68; Conservative 52; Mismatches 119; Indels 70; Gaps 16;

QY 47 SIVFELTLOHSVAVPAE-----EQATGSKSLYYGVKDLATVFFYMLVAIIHATIOE 99
Db 78 AICFACLLSPSLRPFAEPFIFLSYKQPDGS---YKGPADACFPFWVFTAFRVIMD 134
QY 100 VYLDKI-----NKRMQFTKAKONKFNESQFSVFFFCIHW--GTFLISENCISDPT 150
Db 135 YVFRPEVLNMGVNRKVII-----RFCBQG-YSPFYLL-CFWFLGLYIYRSSNYWSNEE 186
QY 151 LTKARPHSMWTFQMKFFYSOLAYNFAHPELYFQKTKK---QDIPRLVYIGLHL---204
Db 187 KLFDYPOQYMSPLPKAYYLQGLFWLQQLVHLHLEQRADHWQMFHAHVTCALILSY 246
QY 205 ---FHTGAYLLYLNHLGLLL-----VLHY--FVELLSHMCGIFYFSDKDYKIGSLWAI 255
Db 247 GFNFLRVGNAILYIFDLSYIILSGGKMLAYLGFKICDYLFGIF-----VASWY 296
QY 256 V--FILGRVLTIVS---VLTGVFHLAGSQNRNPDALTGNNVLAIAKIAVLSSTCTQAY 310
Db 297 SRHYLFSKILRVVVTNAPETIIGGFHL-----DVPNGYIFNKPIYIAFIILLFTLQL-347
QY 311 VTWNLTILW 319
Db 348 ----LIYIW 352

RESULT 7
S30134
hypothetical protein YKL008c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YKL156
C:Species: Saccharomyces cerevisiae
C:Date: 18-Jun-1993 #sequence_revision 18-Jun-1993 #text_change 24-Sep-1999
C:Accession: S30134; S37819; S17017
R:Boyer, J.; Pascolo, S.; Richard, G.F.; Dujon, B.
Yeast 9, 279-287, 1993
A:Title: Sequence of a 7.8 kb segment on the left arm of yeast chromosome XI reveals mammalian UOG-1 gene.
A:Reference number: S30132; MUID:93255906
A:Accession: S30134
A:Molecule type: DNA
A:Residues: 1-418 <BOY>
A:Cross-references: GB:S59773; NID:g300231; PIDN:AAC60549.1; PID:g300232
R:Boyer, J.; Pascolo, S.; Richard, G.F.; Ghazvini, M.; Colleaux, L.; Thierry, A.; Mol. submitted to the Protein Sequence Database, March 1994
A:Reference number: S37813
A:Accession: S37819
A:Molecule type: DNA
A:Residues: 1-418 <BO2>

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27324

R:White, S.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20345

A:Accession: T27324

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-360 <WIL>

A:Cross-references: EMBL:AL032655; PIDN:CAA21723.1; GSPDB:GN00019; CESP:Y6B3B.10

A:Experimental source: clone Y6B3B

C:Genetics:

A:Gene: CESP:Y6B3B.10

A:Map position: 1

A:Introns: 24/3; 58/1; 88/3; 143/1; 205/2; 310/3

Query Match 5.8%; Score 111.5; DB 2; Length 360;
Best Local Similarity 19.7%; Pred. No. 0.041;
Matches 57; Conservative 55; Mismatches 122; Indels 55; Gaps 10;

QY 5 KKSTKNPVLQEFILQNHADIVSCVGMFFLLGLVFEGETAEASIVFLTLQHSVAVPA--- 61

Db 15 RISTANPIVLA-----GLVFESIPHWFRYARLRDPDYSPSSMI 53

QY 62 ABEQATGSKSLYYGYKDLATVFFYMLVAIIHA---TIQEVYLDKINKRMQFTKAKQNK 118

Db 54 SDFKKVSLNSELYTVLIIASIFTPLRYVLOIRLESWTQOHNLYPFAHKVPESFWKLTYY 113

QY 119 FNESQFSVFFESCIGWTFILISENCLSDPLIW-----KARPHSMWTFQMKFFYISOL 173

Db 114 YGVWTFIA-FYFHMV-----DSHDIFNDPLSMWTEWESGGP--KMHQVQVIYAVQS 164

QY 174 AYWFHAFPELYFQTKKODIPQLQVYIGLHLEHITGAYLLYNHL-----GLLLVLHY 227

Db 165 AYIHSIYATLEMDLWRKDSWLMFVH-----HFIALGULFUSYVDNFTLPALVFLHD 218

QY 228 FVELLSHMGCLFFSDEKQKQKIGSLMAIVFILGRVLTIVSVLTGPHL 276

Db 219 NSDALTLEITKLSFLYKKRTNR--QYKYVFLMGNAFAILFAIWIIFRL 265

RESULT 11

AG1051

probable transport protein Sgat sgat [imported] - Salmonella enterica subsp. enterica ser

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001

C:Accession: AG1051

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; PMID:11677608

A:Accession: AG1051

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-476 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD06860.1; PID:gi16505508; GSPDB:GN00176

C:Genetics:

A:Gene: sgat

Query Match 5.7%; Score 108.5; DB 2; Length 476;
Best Local Similarity 22.0%; Pred. No. 0.1;
Matches 73; Conservative 56; Mismatches 136; Indels 67; Gaps 15;

QY 10 NPPVLQSEFILQNHADIVSCVGMFFLLGLVFEGETAEASIVFLTLQHSVAVPAEQA 66

Db 29 NAPLL-----LGIVTCLGYILLKRSVSIIRKTIKTIIGFMLQAGSGILTSTFKP 79

QY 67 TGSKSLYYGVKVD-----LAT-----VFFYMLVAIIHATIQEVYLDKINK 107

Db 80 VVAKNSEVYINGAISDITYASMMATERMGDAYSWSGYAVLLALALN--ICVYLLRRITG 137

QY 108 RMOFTKAKQNKENESQFSV-FYFFS-CIWGTFFIL-----ISENCLSDPTLIWKA 155

Db 138 INTIMLTGHIMFQQAGLIAVSFIYGSWWTIICTAILVSLVWGTISNMATKPT----- 192

QY 156 RPHSMFTQMKFF--YISQAYWFHAFPELYFQTKKODIPQLQVYIGLHLEH--ITGAY 211

Db 193 ---QEVTDCCGSIGHQOQFASWI-AYKVPFLGKKEESVEDLKPGLWNIFHDNIVSTA 248

QY 212 LLYLNHLGLLLLVHYFVELLSHMCGLFYSDEKYQKIGSLMAIVFILGRVLTIVSVLT 271

Db 249 IVMTIFFGAIL--LSFGIDTVQAMAGKVHTVYILOTGFSFAVAIFITIQGVFMFAELS 306

QY 272 VGPHLAGSONRNPDALTGNNVLAAKIAVLSS 303

Db 307 EAFN-GISQRLIPGA-----VLAIIDCAIYS 331

RESULT 12

T20916

hypothetical protein F14F8.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 26-May-2000

C:Accession: T20916

R:Lloyd, C.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19346

A:Accession: T20916

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-331 <WIL>

A:Cross-references: EMBL:Z92782; PIDN:CAB07188.1; GSPDB:GN00023; CESP:F14F8.3

A:Experimental source: clone F14F8

C:Genetics:

A:Gene: CESP:F14F8.3

A:Map position: 5

A:Introns: 55/3; 95/2; 242/3; 260/3

C:Superfamily: Caenorhabditis elegans hypothetical protein K02H11.4

Query Match 5.5%; Score 106.5; DB 2; Length 331;
Best Local Similarity 24.2%; Pred. No. 0.1;
Matches 63; Conservative 36; Mismatches 80; Indels 81; Gaps 15;

QY 69 SKSLYYGVKDLA-----TVFFYMLVAIIHATIQEVYLDKINKRMQFTKA--- 114

Db 4 SHELKEFTIKDLGIYTSMSFLLSIFLVILFIPIFV---ILNRANRRRDEKTAIYP 60

QY 115 KQKNFESQFSVFFYS--CI-----WGTFILISENCLSDPTLIWKA 157

Db 61 ITNYFYKS--LCVYIFSVTCIIITLILYGPNAFTGTNIIIFVFFLIIF-VHV 117

QY 158 HSMWTFQM---KFFYISQAYWFHAFPEL--YFQTKKQDIPQLQVYIGLHLEHITGAYL 212

Db 118 HHLIIFLMAVQR-----LLVFFPSEPLVTFQKT-----TNKIYT-LH-----FL 159

QY 213 LYLNLHGLLLLVHYFVELLSHMCGLFYSDEKYQKIGSLMAIVFILGRVLTIVSVLT 272

Db 160 FILTHLGLLLDWANYQSLSQIPTLYL-----CYIFLNIQITSAILYI 206

QY 273 GF-----HLAGSQNRNP 284

Db 207 PMVMKIRSFHALLASRTYQP 226

RESULT 13

T11411

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - cat mitochondrion (fragment)

C:Species: mitochondrion Felis silvestris catus (domestic cat)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 12-Nov-1999

C:Accession: T11411
R:Lopez, J.V.; Cevario, S.; O'Brien, S.J.
Genomics 33, 229-246, 1996
A:Title: Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome
A:Accession number: 217268; MUID:96301400
A:Reference: T11411
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-459 <LOP>
A:Cross-references: EMBL:U20753; NID:g1098523; PID:g1098533; PIDN:AAC48578.1
A:Experimental source: female adult; isolate FCA-65; lymphocyte; blood
C:Genetics:
A:Genome: mitochondrion
A:Note: ND4
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 4
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 5.5%; Score 106.5; DB 2; Length 459;
Best Local Similarity 21.3%; Pred. No. 0.15;
Matches 51; Conservative 39; Mismatches 82; Indels 67; Gaps 9;

Qy 137 TPILISENCLSDPTLIW-KARPHSMWTQMKFFYISOLAYWFHAFPELYFOKTKQD--- 192
Db 8 TAMLPMTCLSKPNMIWINSTYLSLSISYLNQLGHSNLSLFFSDLSAPLLV 67
Qy 193 -----IPQLVYIGULHFIITGA-----YLLNLNHLGLLLVHYFVELLSHMCGLFY--- 240
Db 68 LTTWLLPLMASQSHLSKEPSRKLYITMLTLQLLIMFTTATELI-----MFYILF 122
Qy 241 -----FSDEKYQKGISLMAIVFILGRVLTIVSVLVGVFHLAQSNRPDAL 287
Db 123 EATLIPTLIITRWGQTERLNAGLYFLFYTLVGLSPLLVALLYI-----QN-----T 170
Qy 288 TGNVNVLAAKIATVLSSTCIQAYVWNLTLW-----LQWVEDSNICA 331
Db 171 TGTNLFLLIQWAKPIS-----TTWSNIFLWACMAFWKMPYGLHLMUPRAHVEA 223

RESULT 14
A86116
hypothetical protein sgat [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: A86116
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A86116
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-484 <STO>
A:Cross-references: GB:AB005174; NID:g12519184; PIDN:AAG59389.1; GSPDB:GN00145; UWGP:Z58
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: sgat

Query Match 5.5%; Score 106.5; DB 2; Length 484;
Best Local Similarity 21.7%; Pred. No. 0.16;
Matches 72; Conservative 56; Mismatches 137; Indels 67; Gaps 14;

Qy 10 NPPVLSQEFILLQNHADIYSCVGMFFL---LGLVFFGTAEASIVELTLOHSVAVPAAEQA 66
Db 37 NAPLL-----LGIYVTLGYYILLRSVSVIIRKTIIGTFMQLQAGSGILSTFKP 87
Qy 67 TGSLSLYYGVKND-----LAT-----VFFYMLVAIIHATIQEYVLDKINK 107
Db 88 VVAKMSEYVINGAISDTYASMMATIDRMGDAYSWGVAVLLALAIN--ICVLLRRITG 145
Qy 108 RMOFTKAKONKFNESGQFSV---FYFFSCIWGTFFL-----ISENCLSDPTLIWKA 155
Db 146 IRTIMLTGHIMFQQAGLIAVTLFIYFGYSWMTTIICTAILVSLYWGITSNMKYKPT----- 200
Qy 156 RPHSMWTQMKFF--YISQLAYWFHAFPELYFOKTKKODIPQLVYIGLHLFH--ITGAY 211
Db 201 ---QEVTDGCGFSIGHQOQFASWI--AYKVAFLPGKKEVEDLKLPGMLNIFHDIVSTA 256
Qy 212 LLYLNHLGLLLVHYFVELLSHMCGLFYFSDERKQKISLMAIVFILGRVLTIVSVLT 271
Db 257 IVMTIFFGAIL--LSFGIDTVQAMAGKVHWTYIILQTGFSFAVAIFITITQGVRMFVAELS 314

Db 146 IRTIMLTGHIMFQQAGLIAVTLFIYFGYSWMTTIICTAILVSLYWGITSNMKYKPT----- 200
Qy 156 RPHSMWTQMKFF--YISQLAYWFHAFPELYFOKTKKODIPQLVYIGLHLFH--ITGAY 211
Db 201 ---QEVTDGCGFSIGHQOQFASWI--AYKVAFLPGKKEVEDLKLPGMLNIFHDIVSTA 256
Qy 212 LLYLNHLGLLLVHYFVELLSHMCGLFYFSDERKQKISLMAIVFILGRVLTIVSVLT 271
Db 257 IVMTIFFGAIL--LSFGIDTVQAMAGKVHWTYIILQTGFSFAVAIFITITQGVRMFVAELS 314
Qy 272 VGFHLAQSNRPDALTCNVNLAAKIATVLS 303
Db 315 EAFN-GISORLIPGA-----VLAIIDCAAIYS 339

RESULT 15
D65230
hypothetical 52.9 kD protein in aidB-rpsF intergenic region - Escherichia coli (strai
N:Alternate names: hypothetical protein o488
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: D65230; S56418
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: D65230
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-484 <BLAT>
A:Cross-references: GB:AB000491; GB:U00096; NID:g2367357; PIDN:AAC77150.1; PID:g23673
A:Experimental source: strain K-12, substrain MG1655
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from
A:Reference number: S56314; MUID:95334362
A:Accession: S56418
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-469, 'AOKKKNNKNSLNKEF' <BUR>
A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97089.1; PID:g537034
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1994
C:Genetics:
A:Gene: YJfs
A:Start codon: GTG

Query Match 5.5%; Score 105.5; DB 2; Length 484;
Best Local Similarity 21.7%; Pred. No. 0.19;
Matches 72; Conservative 56; Mismatches 137; Indels 67; Gaps 14;

Qy 10 NPPVLSQEFILLQNHADIYSCVGMFFL---LGLVFFGTAEASIVELTLOHSVAVPAAEQA 66
Db 37 NAPLL-----LGIYVTLGYYILLRSVSVIIRKTIIGTFMQLQAGSGILSTFKP 87
Qy 67 TGSLSLYYGVKND-----LAT-----VFFYMLVAIIHATIQEYVLDKINK 107
Db 88 VVAKMSEYVINGAISDTYASMMATIDRMGDAYSWGVAVLLALAIN--ICVLLRRITG 145
Qy 108 RMOFTKAKONKFNESGQFSV---FYFFSCIWGTFFL-----ISENCLSDPTLIWKA 155
Db 146 IRTIMLTGHIMFQQAGLIAVTLFIYFGYSWMTTIICTAILVSLYWGITSNMKYKPT----- 200
Qy 156 RPHSMWTQMKFF--YISQLAYWFHAFPELYFOKTKKODIPQLVYIGLHLFH--ITGAY 211
Db 201 ---QEVTDGCGFSIGHQOQFASWI--AYKVAFLPGKKEVEDLKLPGMLNIFHDIVSTA 256
Qy 212 LLYLNHLGLLLVHYFVELLSHMCGLFYFSDERKQKISLMAIVFILGRVLTIVSVLT 271
Db 257 IVMTIFFGAIL--LSFGIDTVQAMAGKVHWTYIILQTGFSFAVAIFITITQGVRMFVAELS 314

Qy 272 VGFHLAGSQNRNPDALTGNVNVLAAKIAYLSS 303
|: || | | || | : |
Db 315 EAFN-GISQRLIPGA-----VLAIDCAAIYS 339

Search completed: September 6, 2002, 17:11:52
Job time: 6791 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 17:21:14 ; Search time 26.42 Seconds
(without alignments)
540.784 Million cell updates/sec

Title: US-09-807-470-4
Perfect score: 1920
Sequence: 1 MGLRKKSTKNPPVLSQEFIL.....VGVTSTNRVDCPPKREKSS 369

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1400.5	72.9	373	1	TRAM_CANFA
2	1387.5	72.3	373	1	TRAM_HUMAN
3	1331.5	69.3	358	1	TRAM_BOVIN
4	832.5	43.4	370	1	Y557_HUMAN
5	144	7.5	411	1	LAGL1_YEAST
6	128	6.7	390	1	LAGL1_SCHPO
7	123.5	6.4	418	1	YKAB8_YEAST
8	116.5	6.1	384	1	YHXF_SCHPO
9	115	6.0	614	1	NUOL_BUCAI
10	106.5	5.5	459	1	NU4M_FELCA
11	105.5	5.5	484	1	SGAT_ECOLI
12	102	5.3	692	1	SEAG_HUMAN
13	99	5.2	802	1	YGN9_YEAST
14	98.5	5.1	391	1	CYB_PART2
15	98.5	5.1	506	1	HITB_HAEIN
16	98.5	5.1	517	1	QOX1_SULAC
17	98	5.1	387	1	YML4_PART2
18	95.5	5.0	368	1	CYB_TOXGO
19	95	4.9	499	1	MYIN_AQUAE
20	94.5	4.9	489	1	YHIP_ECOLI
21	94.5	4.9	696	1	LSHR_PIG
22	93.5	4.9	394	1	YD2C_SCHPO
23	93	4.8	633	1	S6A9_RAT
24	92.5	4.8	538	1	LSHR_SHEEP
25	91	4.7	638	1	YDAG_BOVIN
26	90.5	4.7	531	1	YDFG_SCHPO
27	90	4.7	1564	1	PDRA_YEAST
28	89	4.6	369	1	Y316_MYCPN
29	89	4.6	633	1	S6A9_MOUSE
30	89	4.6	2841	1	NFI_MOUSE
31	88.5	4.6	457	1	YIRO_YEAST
32	88.5	4.6	496	1	Y048_UREPA
33	88	4.6	461	1	SE12_CAEL

34	88	4.6	559	1	INXA_CAEL	Q22549 caenorhabdi
35	88	4.6	836	1	YLAK_CAEL	Q20076 caenorhabdi
36	87.5	4.6	401	1	Y443_MYCPN	P75134 mycoplasma
37	87	4.5	388	1	YF02_AQUAE	O67472 aquifex aeo
38	87	4.5	569	1	HXT8_YEAST	P40886 saccharomyc
39	87	4.5	1529	1	PDRF_YEAST	Q04182 saccharomyc
40	86.5	4.5	491	1	Y225_MYCPN	P75463 mycoplasma
41	86.5	4.5	528	1	TIC1_CHLTR	O84068 chlamydia t
42	86.5	4.5	862	1	TRP7_HUMAN	Q9HCX4 homo sapien
43	86	4.5	2820	1	NFI_RAT	P97526 rattus norv
44	86	4.5	2839	1	NFI_HUMAN	P21359 homo sapien
45	85.5	4.5	254	1	YC8A_METJA	P81318 methanococc

ALIGNMENTS

RESULT 1	TRAM_CANFA	STANDARD;	PRT;	373 AA.
AC	Q01685;			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	TRAM protein (Translocating chain-associating membrane protein).			
OS	Canis familiaris (Dog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
OX	NCBI_TaxID=9615;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-26 AND 164-184.			
RC	TISSUE=Kidney;			
RX	MEDLINE=92244357; PubMed=1315422;			
RA	Goeurlich D., Hartmann E., Prehn S., Rapoport T.A.;			
RT	"A protein of the endoplasmic reticulum involved early in polypeptide translocation.";			
RL	Nature 357:47-52(1992).			
CC	-1- FUNCTION: STIMULATORY OR REQUIRED FOR THE TRANSLOCATION OF SECRETORY PROTEINS ACROSS THE ER MEMBRANE.			
CC	-1- SUBCELLULAR LOCATION: Endoplasmic reticulum membrane.			
CC	-1- SIMILARITY: BELONGS TO THE LASSI FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X63678; CAA45217.1; -			
DR	PIR; S21736; S21736.			
KW	Endoplasmic reticulum; Transmembrane; Glycoprotein; Translocation.			
FT	INIT_MET 0			
FT	DOMAIN 1 28			CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 29 49			POTENTIAL.
FT	DOMAIN 50 75			LUMENAL (POTENTIAL).
FT	TRANSMEM 76 96			POTENTIAL.
FT	DOMAIN 97 120			CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 121 141			POTENTIAL.
FT	DOMAIN 142 158			LUMENAL (POTENTIAL).
FT	TRANSMEM 159 179			POTENTIAL.
FT	DOMAIN 180 191			CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 192 212			POTENTIAL.
FT	DOMAIN 213 216			LUMENAL (POTENTIAL).
FT	TRANSMEM 217 237			POTENTIAL.
FT	DOMAIN 238 250			CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 251 271			POTENTIAL.
FT	DOMAIN 272 296			LUMENAL (POTENTIAL).
FT	TRANSMEM 297 317			POTENTIAL.
FT	DOMAIN 318 373			CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD 55 55			N-LINKED (GLCNAC...)(PROBABLE).
SQ	SEQUENCE 373 AA; 43029 MW; 1D85808E1D80E835 CRC64;			

```
Query Match 72.9%; Score 1400.5; DB 1; Length 373;
Best Local Similarity 72.6%; Pred. No. 7.5e-108;
Matches 270; Conservative 40; Mismatches 57; Indels 5; Gaps 2;

QY 3 LRKSTKNPPVLSQEFILQNHADIVSCVGMFFLLGLVFEQTAEASIVFLQLQHSVAVPAA 62
Db 2 IRKSTKSPVLSHEEFLQNHADIVSCVAMVFLGLMFEITAKASIIFFVTLQYNVTLPAT 61
QY 63 EQATGSKSLYYGVKDLATVFFVLMVAIIHATIQEYVLDKINKRMQFTKAKONKFNES 122
Db 62 EQATGSKSLYYGVKDLATVFFVLMVAIIHATIQEYVLDKINKRMHFSKTHSKFNES 121
QY 123 GQSFVFFFCINGTFFILISENCISDPTLIWKARPHSMFTFQMKFFYISOLAYWFHAFPE 182
Db 122 GQSFVFFFCINGTFFILISENCISDPTLIWKARPHSMFTFQMKFFYISOLAYWFHAFPE 181
QY 183 LYFQTKKQDIPROLYVIGLHFLHITGAYLLYNHLGLLLVLLHYFVELLSHMCGLFYFS 242
Db 182 LYFQTKKQDIPROLYVIGLHFLHITGAYLLYNHLGLLLVLLHYFVELLSHMCGLFYFS 241
QY 243 DEKYOGISLWAVFVLGRVTLVSVLTGFGHLAGSONRNPDLATGNNVLAIAVL 302
Db 242 DEKYOGISLWAVFVLGRVTLVSVLTGFGHLAGSONRNPDLATGNNVLAIAVL 301
QY 303 SSCTIQAVTWNLTLMQWRVEDSNIOASCMKKK----RSRSKKRTENG-VGVTSNR 357
Db 302 SICVQAFMMKKFNFOLRRWRHESAFQAPVKKKPTVTKGRSSKKTENGVTLSNV 361
QY 358 VDCPPKREKSS 369
Db 362 ADSPRNKKKSS 373

RESULT 2
TRAM_HUMAN ID TRAM_HUMAN STANDARD; PRT; 373 AA.
AC Q15629;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TRAM protein (translocating chain-associating membrane protein).
GN TRAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92244357; PubMed=1315422;
RA Goerlich D., Hartmann E., Prehn S., Rapoport T.A.;
RT "A protein of the endoplasmic reticulum involved early in polypeptide translocation."
RL Nature 357:47-52(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: STIMULATORY OR REQUIRED FOR THE TRANSLOCATION OF SECRETORY PROTEINS ACROSS THE ER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum membrane.
CC -!- SIMILARITY: BELONGS TO THE LASSI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC
```

```
DR EMBL; X63679; CAA45218.1; -.
DR EMBL; BC000687; AA000687.1; -.
DR MIM; 605190; -.
KW Endoplasmic reticulum; Transmembrane; Glycoprotein; Translocation.
FT INIT_MET 0 0 BY SIMILARITY.
FT DOMAIN 1 28 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 29 49 POTENTIAL.
FT DOMAIN 50 75 LUMENAL (POTENTIAL).
FT TRANSMEM 76 96 POTENTIAL.
FT DOMAIN 97 120 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 121 141 POTENTIAL.
FT DOMAIN 142 158 LUMENAL (POTENTIAL).
FT TRANSMEM 159 179 POTENTIAL.
FT DOMAIN 180 191 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 192 212 POTENTIAL.
FT DOMAIN 213 216 LUMENAL (POTENTIAL).
FT TRANSMEM 217 237 POTENTIAL.
FT DOMAIN 238 250 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 251 271 POTENTIAL.
FT DOMAIN 272 296 LUMENAL (POTENTIAL).
FT TRANSMEM 297 317 POTENTIAL.
FT DOMAIN 318 373 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 55 55 N-LINKED (GLCNAC... ) (PROBABLE).
SQ SEQUENCE 373 AA; 42940 MW; C220949AF4EEDD0 CRC64;

Query Match 72.3%; Score 1387.5; DB 1; Length 373;
Best Local Similarity 71.5%; Pred. No. 8.8e-107;
Matches 266; Conservative 44; Mismatches 57; Indels 5; Gaps 2;

QY 3 LRKSTKNPPVLSQEFILQNHADIVSCVGMFFLLGLVFEQTAEASIVFLQLQHSVAVPAA 62
Db 2 IRKSTKSPVLSHEEFLQNHADIVSCVAMVFLGLMFEITAKASIIFFVTLQYNVTLPAT 61
QY 63 EQATGSKSLYYGVKDLATVFFVLMVAIIHATIQEYVLDKINKRMQFTKAKONKFNES 122
Db 62 EQATGSKSLYYGVKDLATVFFVLMVAIIHATIQEYVLDKINKRMHFSKTHSKFNES 121
QY 123 GQSFVFFFCINGTFFILISENCISDPTLIWKARPHSMFTFQMKFFYISOLAYWFHAFPE 182
Db 122 GQSFVFFFCINGTFFILISENCISDPTLIWKARPHSMFTFQMKFFYISOLAYWFHAFPE 181
QY 183 LYFQTKKQDIPROLYVIGLHFLHITGAYLLYNHLGLLLVLLHYFVELLSHMCGLFYFS 242
Db 182 LYFQTKKQDIPROLYVIGLHFLHITGAYLLYNHLGLLLVLLHYFVELLSHMCGLFYFS 241
QY 243 DEKYOGISLWAVFVLGRVTLVSVLTGFGHLAGSONRNPDLATGNNVLAIAVL 302
Db 242 DEKYOGISLWAVFVLGRVTLVSVLTGFGHLAGSONRNPDLATGNNVLAIAVL 301
QY 303 SSCTIQAVTWNLTLMQWRVEDSNIOASCMKKK----RSRSKKRTENG-VGVTSNR 357
Db 302 SICVQAFMMKKFNFOLRRWRHESAFQAPVKKKPTVTKGRSSKKTENGVTLSNV 361
QY 358 VDCPPKREKSS 369
Db 362 ADSPRNKKKSS 373

RESULT 3
TRAM_BOVIN ID TRAM_BOVIN STANDARD; PRT; 358 AA.
AC Q9GKZ4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TRAM protein (translocating chain-associating membrane protein) (Fragment).
GN TRAM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
```



```
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Clark T.G., Morris J., Akamatsu M., McGraw R.A., Ivarie R.D.;
RT "Cloning and sequence analysis of a bovine tram CDNA.";
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: STIMULATORY OR REQUIRED FOR THE TRANSLLOCATION OF
CC SECRETORY PROTEINS ACROSS THE ER MEMBRANE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM MEMBRANE (BY
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE LASSI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U19578; AAG10391.1; -.
CC Endoplasmic reticulum; Transmembrane; Glycoprotein; Translocation.
CC NON_TER 1 1
CC DOMAIN <1 13 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 14 34 POTENTIAL.
CC DOMAIN 14 34 POTENTIAL.
CC TRANSMEM 35 60 LUMENAL (POTENTIAL).
CC TRANSMEM 61 81 POTENTIAL.
CC TRANSMEM 82 105 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 106 126 POTENTIAL.
CC TRANSMEM 127 143 LUMENAL (POTENTIAL).
CC TRANSMEM 144 164 POTENTIAL.
CC TRANSMEM 165 176 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 177 197 POTENTIAL.
CC TRANSMEM 198 201 LUMENAL (POTENTIAL).
CC TRANSMEM 202 222 POTENTIAL.
CC TRANSMEM 223 235 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 236 256 POTENTIAL.
CC TRANSMEM 257 281 LUMENAL (POTENTIAL).
CC TRANSMEM 282 302 POTENTIAL.
CC TRANSMEM 303 358 CYTOPLASMIC (POTENTIAL).
CC CARBOHYD 40 40 N-LINKED (GLCNAC... (PROBABLE).
CC SEQUENCE 358 AA; 41403 MW; 25CF9930C4CDDA15 CRC64;
Query Match 69.3%; Score 1331.5; DB 1; Length 358;
Best Local Similarity 71.8%; Pred. No. 3.3e-102;
Matches 257; Conservative 37; Mismatches 59; Indels 5; Gaps 2;
QY 17 EFILQNHADIVSCVGMFFLLGLVFEPTAEASIVFLTLQHSVAVPAABEQATGSKSLYYG 76
DB 1 EFVLQNHADIVSCVAMVFLGLMFEITAKVSIIFVTLLQNVNLTLPATEEQATESAFLLYYG 60
QY 77 VKDLATVFFVMLVAIIHATIQEYVLDKINKRQFTKAKONKNEQSFVFFFCISWG 136
DB 61 IKDLATVFFVMLVAIIHATIQEYVLDKINKRHFHSKTHSKFNEQSLSAFLFCISWG 120
QY 137 TFLISENCLSDPTLWKARPHSMFTOMKFFVYISOLAYWFHAPPELYFOKTKKQDIPRO 196
DB 121 TFLISENYSIDPTLWRAVPHNLTOMKFFVYISOLAYWFHAPPELYFOKTKKQDIPRO 180
QY 197 LVYIGLHLFHTGAYLLYLNLHGLLLLVHYFVELLSHMCGLFYFDEKQKQISLWAV 256
DB 181 LVYIGLFLPHIAGAYLLNLHGLLVLLVHYFVEFLFHSRLSLEYFDEKQKQISLWAVL 240
QY 257 FIIGRLVTLIVSVLTGVLHAGSONRNPDALTGNNVNLAAKIAIVLSSCTQIAYVTWNLI 316
DB 241 FVLGRLLTLIVSVLTGVLHAGLARAENQKLDFTGNFNVLAIVRAVLASICITQAFMMWKEI 300
QY 317 TLWLQWRVDSNTQASCKKK-----RSRSKKRTKENGCV-GVETSNRVDCPPKREKSS 369
DB 301 NFQLRRRREHSATQAPAVKKPKPVTKGRSXXKGTENGVTGTVTNGADSPRNRKREKSS 358
```

```
RESULT 4
Y557 HUMAN STANDARD; PRT; 370 AA.
AC Q15035;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0057.
GN KIAA0057.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE-Bone marrow; PubMed=7584044;
RX MEDLINE=96051398;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 1:223-229(1994).
RN [2]
RP SEQUENCE FROM N.A.
RP Tracey A.;
RA Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE LASSI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D31762; BAA06540.1; -.
CC EMBL; AL049611; CAB71119.1; -.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 23 43 POTENTIAL.
FT TRANSMEM 76 96 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 251 271 POTENTIAL.
FT TRANSMEM 288 308 POTENTIAL.
SQ SEQUENCE 370 AA; 43327 MW; 9B5183F1A3D45366 CRC64;
Query Match 43.4%; Score 832.5; DB 1; Length 370;
Best Local Similarity 45.1%; Pred. No. 3.1e-61;
Matches 171; Conservative 72; Mismatches 115; Indels 21; Gaps 7;
QY 1 MGLRKKSTKNPPVLSQEFILQNHADIVSCVGMFFLLGLVFEPTAEASIVFLTLQHSVAVP 60
DB 1 MAFRRR-TKSYPLSFQEFVHNHADIGFCLVLCVGLMFEVTAFTFLFIPQYNISVP 59
QY 61 AABEQATGSKSLYYGVKDLATVFFVMLVAIIHATIQEYVLDKINKRQFTKAKONKFN 120
DB 60 TADSETVH----YHYGPKDLVTILFYITFIILHAVVQEVLDKISRLHLSKVHKSFN 115
QY 121 ESQSFVYFFFCISWGTFFILISENCLSDPTLWKARPHSMFTOMKFFVYISOLAYWFHAF 180
DB 116 ESGQVVFHFHTSVTCVFFVVTGTYLTPNPSLWEDYVILKISRLHLSKVHKSFN 175
QY 181 PELYFOKTKKQDIPROLYVIGLHLFHTGAYLLYLNLHGLLLLVHYFVELLSHMCGLFY 240
DB 176 PELYFOKVRKEEIPROLYVIGLHLFHTGAYLLYLNLHGLLLLVHYFVELLSHMCGLFY 235
QY 241 FSDKQYKGLSLWAVIFILGRVLTIVSVLTGVLHAGSONRNPDALTGNNVNLAAKIAV 300
```

Db 236 FADENNEKLSAWAFAVGVTRFILTLAVLAIGFGLARMENQAFDPEKGFNTFLCRLCV 295
QY 301 LSSCTIOAYVTWNLTLWLORWEDSNQASCKKKR-----SRSSKRT---ENG 349
Db 296 LLLVCAAQAWLWRETHSOLRWETNEQSA---KRRVPATPRLPARLIKRESGYHENG 352
QY 350 VGVETSNRVDCPPKRRKS 368
Db 353 V-VKAENGTS-PRTKLKS 369

RESULT 5
LAG1 YEAST
ID LAG1 YEAST STANDARD; PRT; 411 AA.
AC P38703;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Longevity-assurance protein 1 (Longevity assurance factor 1).
GN LAG1 OR YHL003C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=94253121; PubMed=8195187;
RA D'Mello N.P., Childress A.M., Franklin D.S., Kale S.P.,
RA Pinwasdi C., Jazwinski S.M.,
RT "Cloning and characterization of LAG1, a longevity-assurance gene in
RT yeast";
RL J. Biol. Chem. 269:15451-15459(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Lateille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasikis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII";
RL Science 265:2077-2082(1994).
CC -!- FUNCTION: INVOLVED IN THE AGING PROCESS. DELETION OF LAG1 RESULTS
CC IN A PRONOUNCED INCREASE (APPROXIMATELY 50%) IN MEAN AND IN
CC MAXIMUM LIFE SPAN.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE LASS1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U08133; AAA21579.1; -;
DR EMBL; U10555; AAB68429.1; -;
DR PIR; S46800; S46800.
DR SGD; S0000595; LAG1.
KW Transmembrane.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 212 232 POTENTIAL.
FT TRANSMEM 252 272 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 356 376 POTENTIAL.

FT CONFLICT 173 174 ML -> IV (IN REF. 2).
FT CONFLICT 220 220 F -> C (IN REF. 2).
FT CONFLICT 301 411 VFTPEVGLFVFFWLYLRHVNVNIRILWSVLTFRHEGNNVL
FT NPATQYKWCISLPVFLVLAALQVLYWFLILRLYRL
FT IWOGLQKDRSDSDSAENESREKE -> TEISGIWE
FT KQEDSDNDPRTALSPNETSKQVDPDLLVNLPTENRAL
FT LEAIKSRPTTIAIDTSEPSLVTPYIPGNDSDLSRVNPLL
FT GVLARAGQGLQNLARNNEK (IN REF. 1).
SQ SEQUENCE 411 AA; 48454 MW; 91676D56AC053F3C CRC64;

Query Match 7.5%; Score 144; DB 1; Length 411;
Best Local Similarity 22.6%; Pred. No. 0.00011;
Matches 65; Conservative 49; Mismatches 112; Indels 62; Gaps 10;

QY 33 FFLGLVPEGTAEASIVFLTQHSVAVP-----AAEEQATGSKSLYYGVKDLATVFFFM 87
Db 88 FFLVLCVY-----SAYFLSGNRSTESNPLHMFVAISYQVGDGDS-YAKGIKDLSPVFFYM 140
QY 88 LVAILIHATIQEVLDKINKRMQFTKAKONKFESGQSFVYF-FSCIWGTFLI-SENC 145
Db 141 IFFTELRFLMDVIRPFTVYLVNTSEHRQKRMLOMYAIFYCGVSGPGLIMYHSDLM 200
QY 146 LSDPTLINKARPHSMFTQMKFFYISQLAYWPHAPPELYFQTKKODIPRLVYIGL-HL 204
Db 201 LFKTKPMYRTYPTVITNPFLFKFYLGQAFAWAQACVLVLQLEKPKDYKELVFHHIVTL 260
QY 205 FHITGAYLLYLNHLGLLLV-----LHYFVELLS-HMCGLFY----- 240
Db 261 LLIISSYVFHFTKMGILAIYIMDVSDFFLSLSKTLNLSNFTVPFVGLVFFWFIYLRHV 320
QY 241 -----FSDEKYQKIGSLWAIVFILGRVLTIL 266
Db 321 VNIRILWSVLTFRHEGNNVLFNATQYKWCISL-PIVEFLAALQV 367

RESULT 6
LAG1 SCHPO
ID LAG1 SCHPO STANDARD; PRT; 390 AA.
AC P78970; O13860;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Longevity-assurance protein 1 (Longevity assurance factor 1).
GN LAG1 OR SPAC1A6.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Chanda E.R., Lingner C., Ko Z., Young P.G.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V., Wood V.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE AGING PROCESS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE LASS1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U76608; AAB19113.1; -;
DR EMBL; Z99258; CAB16359.1; -;

```
KW Transmembrane.
FT DOMAIN 13 16 POLY-SER.
FT TRANSMEM 63 83 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 231 251 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 336 356 POTENTIAL.
FT TRANSMEM 379 390 EDEASSTNEDK -> GREGGRENE (IN REF. 1).
FT CONFLICT 379 390
SQ SEQUENCE 390 AA; 45668 MW; DC00FB5C2D2F22CC CRC64;

Query Match 6.7%; Score 128; DB 1; Length 390;
Best Local Similarity 22.0%; Pred. No. 0.0022;
Matches 68; Conservative 52; Mismatches 119; Indels 70; Gaps 16;

QY 47 SIVFLQLQSHVAVPAE-----EQATGSKSLYYGVKDLATVFFYMLVAIIHATIQE 99
DB 78 AICFACLLSPSLRPAEPIFLSYKQPDGS---YGRKPKDACPFIFWIVFTAFRVIMVD 134
QY 100 YVLDKI-----NKRMQTKAKQKNFESGQSFVFFSCIW--GTFILISENCLSDPT 150
DB 135 YVPRPVLNMGVNRKVII-----RFEQSG-YSFYYL-CFVFLGLIYRSNYSNNEE 186
QY 151 LIWKARPHSMVTFQMKFFVYSQLAYWFHAFPELYFQTKK---QDIPROLVYIGLHL--- 204
DB 187 KLEFEDPYQYMSPLFRAYLYLQGLFWLQILYLHLQRRADHWQMFHAHIVTCALILSY 246
QY 205 ---FHTGAYLYNLHGLLLL-----VLHY--FVELLSHMCGLFYSDEKYQKGISLWAI 255
DB 247 GFNFLRVGNAILYFDLSYLLGSGKMLYLGKICDYLFGIF-----VASWVY 296
QY 256 V--FILGRVLTIVS---VLTGVFHLGASQNRNPDALTGNNVNLAKIAVLSSCTIQAY 310
DB 297 SRHYLSKILRVVVTNAPETIIGFHL-----DVPNGYIFNKPIYIAFIILLETQL- 347
QY 311 VTWNLTWL 319
DB 348 ----LIIW 352

RESULT 7
YKAB YEAST
ID YKAB YEAST STANDARD; PRT; 418 AA.
AC P28496;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 49.0 kDa protein in UFD4-CAP1 intergenic region.
GN YKL008C OR YKL156.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93255906; PubMed=8488728;
RA Boyer J., Pascolo S., Richard G.F., Dujon B.;
RT "Sequence of a 7.8 kb segment on the left arm of yeast chromosome XI
RT reveals four open reading frames, including the CAP1 gene, an intron-
RT containing gene and a gene encoding a homolog to the mammalian UOG-1
RT gene."
RL Yeast 9:279-287(1993).
RN [2]
RP SEQUENCE OF 1-149 FROM N.A.
RX MEDLINE=93077675; PubMed=1447293;
RA Amatruda J.F., Gattermeir D.J., Karpova T.S., Cooper J.A.;
RT "Effects of null mutations and overexpression of capping protein on
RT morphogenesis, actin distribution and polarized secretion in yeast."
RL J. Cell Biol. 119:1151-1162(1992).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE LASS1 FAMILY.
```

```
CC -----SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X61398; CAA43670.1; -
DR EMBL; S59773; AAC60549.1; -
DR EMBL; Z28008; CAA81843.1; -
DR PIR; S30134; S30134.
DR SGD; S0001491; YKL008C.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 173 193 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 356 376 POTENTIAL.
SQ SEQUENCE 418 AA; 48992 MW; 7691BA623AC0460A CRC64;

Query Match 6.4%; Score 123.5; DB 1; Length 418;
Best Local Similarity 19.4%; Pred. No. 0.0056;
Matches 59; Conservative 59; Mismatches 117; Indels 69; Gaps 12;

QY 6 KSTKNPVLISQEF-----ILQNHADI-----VSCVMFFLGLVFEGETAEASIV--FL 51
DB 59 EATKNDSDLVKKIWFISFREISYRHWIAPLMILIAVYSAYFTSG---NTTKTNVLRHV 114
QY 52 TLOHSAVVAFAAEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINRQOF 111
DB 115 AVSYQI-----GDTNAYGKGINDLFCFYIMIFFTFLREFLMDVIVRFAIRLHV 164
QY 112 TRAKONKFNESGQSFVY-----FFSCIWGTFLISENCLSDPTLIWKARPHS 159
DB 165 TSKHRIKRIEMQYAIPTYGVSGPFGIYCMYHSDLW-----FFNTKAMRYYPDF 214
QY 160 MMTFQMKFFYISQLAYWFHAFPELYFQTKKQDIPROLVYIGL-HLFHITGAYLLYNHL 218
DB 215 TNPFLKVFYLGQAFAWQAQACILVLQLEKPRKDHNLTFHIVTLLLIWSSVVFHTKM 274
QY 219 GLLLV-----LHYFVELLSHM-CGLFYFSDKQKGISLWAIIFILGR---LTVLIV 267
DB 275 GLPIYITMDVSDFLLSFKTLNLYDSGLAFFS-----FAIEVVAWIYLRHYINLKILW 327
QY 268 SVLT 271
DB 328 SVLT 331

RESULT 8
YHXF SCHPO
ID YHXF SCHPO STANDARD; PRT; 384 AA.
AC O59735;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 42.4 kDa protein C3E17.15c in chromosome II.
GN SPBC3E7.15C OR SPBC4F6.02C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lyne M., Wood V., Rajandream M.A., Barrell B.G., Brown D.,
RA Churcher C.M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
```

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Gwilliam R., Rajandream M.A., Barrell B.G., Skellon J., Churcher C.M.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE LASSI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL023534; CAAL9018.2; -
CC EMBL; AL031534; CA20722.2; -
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 195 215 POTENTIAL.
FT TRANSMEM 233 253 POTENTIAL.
FT TRANSMEM 270 290 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
SQ SEQUENCE 384 AA; 45335 MW; 82FCF8EA6638849A CRC64;

Query Match 6.18; Score 116.5; DB 1; Length 384;
Best Local Similarity 21.68; Pred. No. 0.019;
Matches 50; Conservative 41; Mismatches 121; Indels 19; Gaps 8;

Qy 2 GLRKKSTKNPPVLSQEFILQNHAD--IVSCVGMFLLGLVF---EGTAEASIVFTLQHS 56
Db 34 GRRRSKISVGRAGNAVLRSKERTWIPLILLLVGVFWPNNGYIKYGI-FL----S 88

Qy 57 VAVPAEEQATGSKSLYGYGVKDLATVFFYMLVAIIHATIOEYVLDKINKRMQF-TKAK 115
Db 89 YPIPGT-----NPAQYKGRGLDIAFLYALFTFCFEFTMQEIIARIGRHFNIRAPAK 142

Qy 116 QKNFESQSFVFFFCISGWTFL-ISENCLSDPTLLWKARPHSMFTQMKFFYISOLA 174
Db 143 LRRFEQYATCLYFTVMGSLYVNMKQPMWFFNTDAFWEEYHPHYHVGSFKAFYLIEAA 202

Qy 175 YWFHAPPELYFQKTK-KODIPROLVYIGLHLFHTGAYLLYLNLHLLLV 224
Db 203 YWIQALVLLIQLEKPRDFELVHVHIIITLLIGLSYFHTWIGLAVFI 253

RESULT 9
NUOL_BUCAI STANDARD; PRT; 614 AA.
AC P57262;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE NADH dehydrogenase I chain L (EC 1.6.5.3) (NADH-ubiquinone
DE oxidoreductase chain L).
GN NUOL OR BU164.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10930377;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
```

```

CC -!- SUBUNIT: COMPOSED OF 13 DIFFERENT SUBUNITS. SUBUNITS NUOA, H, J,
CC K, L, M, N CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC OF CHLOROPLASTS OR MITOCHONDRIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP001118; BAB12882.1; -
CC InterPro; IPR001750; Oxidored_q1
CC InterPro; IPR001516; Oxidored_q1_N.
CC Pfam; PF00361; oxidored_q1; 1.
CC Oxidoreductase; NAD; Ubiquinone; Transmembrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 136 156 POTENTIAL.
FT TRANSMEM 168 188 POTENTIAL.
FT TRANSMEM 207 227 POTENTIAL.
FT TRANSMEM 247 267 POTENTIAL.
FT TRANSMEM 271 291 POTENTIAL.
FT TRANSMEM 327 347 POTENTIAL.
FT TRANSMEM 372 392 POTENTIAL.
FT TRANSMEM 410 430 POTENTIAL.
FT TRANSMEM 455 475 POTENTIAL.
FT TRANSMEM 492 512 POTENTIAL.
FT TRANSMEM 533 553 POTENTIAL.
FT TRANSMEM 593 613 POTENTIAL.
SQ SEQUENCE 614 AA; 70881 MW; 18CCC2DFC4FE27E0 CRC64;

Query Match 6.0%; Score 115; DB 1; Length 614;
Best Local Similarity 20.28; Pred. No. 0.043;
Matches 59; Conservative 52; Mismatches 91; Indels 90; Gaps 14;

Qy 80 LATVFF---YMLVAIIHATIOEYVLDKINKRMQFTKAKQKNFESQSFV--YFFSCI 134
Db 1 MSIIFFILPLIGFLSTIODTFKRYTLNI-----GIFSIFISFTICF 47

Qy 135 WGTILISENCLSDPTLLWKARPHSMFTQMKF-----FYISQLAYWHPAPE 182
Db 48 YGVSILKNNNQVF-TQILWKWL--SINEFKIDFGFFLDGLSLMLFVITGVGLLIHFSS 104

Qy 183 LYFQTKKODIPROLVYIGLHLFHTGAYL-----LYLNHLGL----LLVLVHYFVELL 232
Db 105 WMYRKEQS--RFAYTNLFIASMSVLVLADNLFMTLWEGVSVCSYLLIGFYITELK 162

Qy 233 SHMGGLFYFSDKYQKGISLWAIIFLGR--VTLLIVSLVTVGFHLAGSQNRNPDALTN 290
Db 163 NNLCAF-----KAFILTRVSDVFLMIGMELI-YREFNSFN----- 196

Qy 291 VNVLAATVAVLSSTCTQAYVTWNLITLW-----LQRWVEDSNI 329
Db 197 ----FQEIKFLLSSFLNVENFYLDYITLILLGVIGKSAQLPLQTLSDAMV 244

RESULT 10
NU4M_FELCA STANDARD; PRT; 459 AA.
ID NU4M_FELCA
AC P48916;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3).
GN MTND4 OR ND4.
```

```
OS Felis silvestris catus (Cat).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=96301400; PubMed=8660972;
RA Lopez J.V., Cevario S., O'Brien S.J.;
RT "Complete nucleotide sequences of the domestic cat (Felis catus)
RT mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in
RT the nuclear genome.";
RL Genomics 33:229-246(1996).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U20753; AAC48578.1; -.
DR InterPro; IPR003918; NADHub_oxdrdctse4.
DR InterPro; IPR001750; Oxidored_q1.
DR InterPro; IPR000260; Oxidored_q5_N.
DR Pfam; PF00361; oxidored_q1; 1.
DR Pfam; PF01059; oxidored_q5_N; 1.
DR PRINTS; PR01437; NUOXDRDTASE4.
DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
KW OXIDOREDUCTASE; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 459 AA; 51497 MW; C1FA567606F1BFB2 CRC64;

Query Match 5.5%; Score 106.5; DB 1; Length 459;
Best Local Similarity 21.3%; Pred. No. 0.15;
Matches 51; Conservative 39; Mismatches 82; Indels 67; Gaps 9;

QY 137 TETLISENCLSDPTLW-KARPHSMFTQKFFYISQLAWFHAFPELYFQKTKQD--- 192
Db 8 TAMDMPKLSKPNMIWINSTTYSLSLSLSLNLGSHSNFSLFFSDSLSLAPLV 67

QY 193 -----IPRQVYIGLHLFHTGA-----YLLYNHLGLLLLVHYFVLLSHMGLFV--- 240
Db 68 LTTWLLPLMLASQSHLSKETPSRKLYITMLTLQLLLIMTFTATELI-----MEYILF 122

QY 241 -----FSDEKYQKGISLNAIVPILGRVLTVLIVSVLTGVFHLAGSONRNPDAL 287
Db 123 EATLIPTLIITRWGQDTERLNAGLYFLFVTLVGLPLVALLYI-----QN-----T 170

QY 288 TGNVNVYLAAKIAYLSSCTIQAVVTWNLTW-----LQRWVEDSNIOA 331
Db 171 TGTINFLIIQYWKAPIS-----TTSNIFLWLAACMAFMVKMPLXGLHLWLPKRAHVEA 223

RESULT 11
SGAT_ECOLI
ID SGAT_ECOLI STANDARD; PRT; 484 AA.
AC P39301;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative transport protein sgat.
GN SGAT OR B4193.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
```

```
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP DISCUSSION OF SEQUENCE.
RA Reizer J., Charbit A., Reizer A., Saier M.H. Jr.;
RT "Novel phosphotransferases system genes revealed by bacterial genome
RT analysis: operons encoding homologues of sugar-specific permease
RT domains of the phosphotransferase system and pentose catabolic
RT enzymes.";
RL Genome Sci. Technol. 1:53-75(1996).
CC -!- FUNCTION: COULD ACT AS THE TRANSPORT PROTEIN FOR THE UNKNOWN
CC PENTITOL SUBSTRATE OF THE SGA OPERON.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: TO M.PNEUMONIAE SGAT HOMOLOG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14003; AAA97089.1; -.
DR EMBL; AE000491; AAC77150.1; -.
DR EcoGene; EG12493; sgat.
KW Transport; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 57 77 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
FT TRANSMEM 252 272 POTENTIAL.
FT TRANSMEM 282 302 POTENTIAL.
FT TRANSMEM 335 355 POTENTIAL.
FT TRANSMEM 357 377 POTENTIAL.
FT TRANSMEM 398 418 POTENTIAL.
FT TRANSMEM 420 440 POTENTIAL.
FT TRANSMEM 446 466 POTENTIAL.
FT CONFLICT 470 484 RAEEDAEKQLAQSQA -> AQKKMKNNQNSLLNKEF
FT (IN REF. 1).
SQ SEQUENCE 484 AA; 52950 MW; 2CBBED0044BC6CAC CRC64;

Query Match 5.5%; Score 105.5; DB 1; Length 484;
Best Local Similarity 21.7%; Pred. No. 0.2;
Matches 72; Conservative 56; Mismatches 137; Indels 67; Gaps 14;

QY 10 NPPVLSQEFILQNHADIVSCVGMFFEL---LGLVFEGTAASIVFLTLQHSVAVPAABEQA 66
Db 37 NAPL-----LGIVTCLGYILLRKSVSVIIKTIKIYIGFMLLQAGSGILTSTFKP 87

QY 67 TGSKSLYYGVKD-----LAT-----VFFYMLVAIIIIHATIQEYVLDKINK 107
Db 88 VYAKMSEVVGINGAISDTYASMMATIDRMGDAYSWGVAYVLLALNLN--ICVLLRRITG 145

QY 108 RMOFTKAKONKNESQFSV--FYFFSCIWGTFIL-----ISENCLSDPTLIWKA 155
Db 146 IRTIMLTGHIMFQAGLIAVTLTFIFGYSMTTICTAILVSLVWGTSTNNMYKPT----- 200
```

QY 156 RPHSMFTQKFF--YISOLAWFHAFPELYFOKTKKODIPQLVYIGLHLFH--ITGAY 211
DB 201 ---QEVTDGCGFSIGHQOQFASWI--AYKAPPLGKKEESVEDLKLPGWLNIPHDNIVSTA 256
QY 212 LLYLNLHGLLLVLYHYVELLSHMCGLFFSDEKQKGLSWAIVFLLGRULVTLVSVLT 271
DB 257 IYMTIFFGAIL--LSFGIDTVQAMQVHMTVYILOTGFSFAVAIFITQGVMEFVAELS 314
QY 272 VGFHLGASNNRPDALTGNVNVLAAKIIVLSS 303
DB 315 EAFN-GISORLIPGA-----VLAIDCAIYS 339
RESULT 12
ID S6A9 HUMAN STANDARD; PRT; 692 AA.
AC P48067;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sodium- and chloride-dependent glycine transporter 1 (GlyT1) (GlyT-1).
GN SLC6A9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94239375; PubMed=8183239;
RA Kim K.-M., Kingsmore S.F., Han H., Yang-Feng T.L., Godinot N.,
RA Seldin M.F., Caron M.G., Giros B.;
RT "Cloning of the human glycine transporter type 1: molecular and
RT pharmacological characterization of novel isoform variants and
RT chromosomal localization of the gene in the human and mouse
RT genomes.";
RL Mol. Pharmacol. 45:608-617(1994).
CC !- FUNCTION: Terminates the action of glycine by its high affinity
CC sodium-dependent reuptake into presynaptic terminals. May play a
CC role in regulation of glycine levels in NMDA receptor-mediated
CC neurotransmission.
CC !- SUBCELLULAR LOCATION: Integral membrane protein.
CC !- ALTERNATIVE PRODUCTS: 3 ISOFORMS; GLYT-1A, GLYT-1B AND GLYT-
CC 1C (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC !- TISSUE SPECIFICITY: ALL GLYT-1 SUBTYPES CAN BE FOUND IN BRAIN,
CC KIDNEY, PANCREAS, LUNG, PLACENTA, LIVER. BUT GLYT-1C SUBTYPE IS
CC ONLY FOUND IN THE BRAIN.
CC !- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNF).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S70609; AAB30784.1; -;
CC DR EMBL; S70612; AAB30785.1; -;
CC DR MIM; 601019; -;
CC DR InterPro; IPR000175; Na_neurotran_symport.
CC DR Pfam; PF00209; SNF; 1.
CC DR PRINTS; PR00176; NANEUSMPORT.
CC DR PROSITE; PS00610; NA_NEUROTRAN_SYMPT_1; 1.
CC DR PROSITE; PS00754; NA_NEUROTRAN_SYMPT_2; 1.
CC DR PROSITE; PS0267; NA_NEUROTRAN_SYMPT_3; 1.
CC DR Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;
CC Symport; Amino-acid transport; Transport; Alternative splicing.
CC FT DOMAIN 1 94 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 95 115 1 (POTENTIAL).
FT TRANSMEM 123 142 2 (POTENTIAL).

FT TRANSMEM 166 186 3 (POTENTIAL).
FT DOMAIN 187 273 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 274 292 4 (POTENTIAL).
FT TRANSMEM 301 318 5 (POTENTIAL).
FT TRANSMEM 354 371 6 (POTENTIAL).
FT TRANSMEM 383 404 7 (POTENTIAL).
FT TRANSMEM 437 456 8 (POTENTIAL).
FT TRANSMEM 485 503 9 (POTENTIAL).
FT TRANSMEM 519 539 10 (POTENTIAL).
FT TRANSMEM 560 579 11 (POTENTIAL).
FT TRANSMEM 598 616 12 (POTENTIAL).
FT DOMAIN 618 692 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 226 226 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 69 MAAAHGPVAPSSPSQVTLTPVQRSEFLPFPFSGATPSTSLAE
FT SVLKVWHGAYNSGLLPQLMAQHSLSMAQ -> MYGKGAKGM
FT L (IN ISOFORM GLYT-1A).
FT VARSPLIC 16 69 MISSING (IN ISOFORM GLYT-1B).
SQ SEQUENCE 692 AA; 76823 MW; FABA3243A0D98073 CRC64;
Query Match 5.3%; Score 102; DB 1; Length 692;
Best Local Similarity 19.3%; Pred. No. 0.57;
Matches 64; Conservative 42; Mismatches 91; Indels 134; Gaps 14;
QY 26 IVSCVGMFFLLGLVFEQTAESIVFLTQHSVAVPAEAEQATGSKSLYYGVKDLATVFF 85
DB 276 LLGCLGVSWL-----VFELCLRGV-----KSSGVVVF-----TATPPY 310
QY 86 YMLVAIIHATIQEYVLDKRM--QFTKAKQKNFESGQSFVFFSCIWGTFILIS- 142
DB 311 VVLTILFVRGVTLEGAFDGMVYLPQWDKILEAKVWGDAASQIEFSLACAMGLITMAS 370
QY 143 -----ENCLSDPTLWKARPHSMWTFQMKFFVISOLATWFHAFPELYFOKTKKQDIPROL 197
DB 371 YNKFHNCRYSVII-----SITNCA-----TS 393
QY 198 VYIGLHLPHITGAYLLYLNLHGLLLVLYHYVELLSHMCGLFFSDEKQKGIS----- 251
DB 394 VYAGVVFISILG---FMANHLGVDV-----SRVADHGFGLAFVA---YPEALTLLPISP 441
QY 252 LMAIYVF-----ILGRVLTLIV-----SVLTGFGHLGASNNRPD 285
DB 442 LWSLFFEFMLILLGLGTQFCLETLVTAIVDEGVNWILOKKTYYTLGVAVAG----- 494
QY 286 ALTGNNVNVLAAKIIVLSSCTIQAVVTWNLI 316
DB 495 -----FLLGIPLTSQAGIYWILL 512
RESULT 13
YGN9_YEAST
ID YGN9_YEAST STANDARD; PRT; 802 AA.
AC P53121;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Hypothetical 90.8 kDa protein in HUL5-SEC27 intergenic region.
GN YGL139W OR G2812.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-749 FROM N.A.
RC STRAIN=S288C / FY1769;
RX MEDLINE=97197983; PubMed=9046099;
RA Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;
RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm
RT chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,
RT TPI1, MRF1 genes and six new open reading frames.";

```
RL Yeast 13:177-182(1997).
RN [2]
RP SEQUENCE OF 616-802 FROM N.A.
RC STRAIN=S288C / FY1679;
RA MEDLINE=96437978; PubMed=8840506;
RX Escribano V., Eraso P., Portillo F., Mazon M.J.;
RT "Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces
   cerevisiae chromosome VII reveals SPC27, SSMlb, a putative
   S-adenosylmethionine-dependent enzyme and six new open reading
   frames."
RT Yeast 12:887-892(1996).
RL Yeast 12:887-892(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO YEAST YAL053W AND S.POMBE SPAC1F7.03.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X99960; CAA68223.1; -
CC EMBL; Z72661; CAA96851.1; -
CC EMBL; Z72660; CAA96850.1; -
CC EMBL; X92670; CAA63357.1; -
CC SGD; S0003107; XGL139W,
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 170 190 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 406 426 POTENTIAL.
FT TRANSMEM 468 488 POTENTIAL.
FT TRANSMEM 496 516 POTENTIAL.
FT TRANSMEM 526 546 POTENTIAL.
FT TRANSMEM 558 578 POTENTIAL.
SQ SEQUENCE 802 AA; 90761 MW; 7BA13714AD912295 CRC64;

Query Match 5.28; Score 99; DB 1; Length 802;
Best Local Similarity 19.38; Pred. No. 1.2;
Matches 79; Conservative 53; Mismatches 134; Indels 144; Gaps 16;

Qy 19 ILQHADVSCVGMFFL-----LGLVPE-----GTEA----- 46
Db 146 ITNNVSEYACIQIYFSGKTVSGYKWTAVVAGIGLLLSAILSTAFGNSAASHISAN 205

Qy 47 -----STVFTLQHSVAPV-----ABE-----QATGSKS 71
Db 206 TMSLFYFQSVVYVAMQHVHPPIAAWAENLVMSGLIRISFMRFRWYVQSTGTP 265

Qy 72 LYYGVKDLATVF-----FYMVAIIHATIQEY-----VLDKINKRMQFTKAKQKENE 121
Db 266 SLYTSTSMVLAORSQWYLMELPLIKRATNVLYGNANTLIFRGIKRYKMGIENTSIV 325

Qy 122 SQQSVFFVFCISGTFILISENCISDPTLI-----WKARPHSMWTFQ---MKFFYI 170
Db 326 CTGTFVFLCGVVLAVFIIVKCCVELATRLGWTQKARFWEFRKQWRMLKGLLRIYI 385

Qy 171 SOLAYWFHAFPELYFQTKKODIPQLVYIGLHLFHITGAYL-----LYL 215
Db 386 G-----FVQLTILSFWEFTERDSPAIVIACLFILLSCGLMLWAARVTFPARRSVALYN 440

Qy 216 NHLGLLLVLYFVLLSHMCGLFY--FSDEKYQKGLSLWAIIVILGRVTLIVSVLTVG 273
Db 441 NPAALL-----YGDEYVLHKYGFYTFNANHYNWNIIVLLSYIF-----VKSLVVG 486

Qy 274 FHLAGSNR-----NP---DALTGNNVLAAKIAVLSS 303
Db 487 FAQASGQTVLFMFILDLFYFVAIIYKPYLDPRPTNIMILIAIVTVVNS 536

RESULT 14
```

```
CYB_PARTE STANDARD; PRT; 391 AA.
ID CYB_PARTE
AC P15585;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cytochrome B.
GN COB OR CVTB.
OS Paramecium tetraurelia.
OG Mitochondrion.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramecium.
OX NCBI_TaxID=5888;
RN [1]
RC STRAIN=STOCK 51;
RX MEDLINE=90174913; PubMed=2308823;
RX Pritchard A.E., Sellhamer J.J., Mahalingam R., Sable C.L.,
   Venuti S.E., Cummings D.J.;
RT "Nucleotide sequence of the mitochondrial genome of Paramecium."
RL Nucleic Acids Res. 18:173-180(1990).
CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X15917; CAA34052.1; -
CC PIR; S07743; S07743.
CC InterPro; IPR000179; Cyt_b_b6.
CC Pfam; PF00033; cytochrome_b_n; 1.
CC PROSITE; PS00192; CYTOCHROME_B_HEME; FALSE NEG.
CC PROSITE; PS00193; CYTOCHROME_B_OO; FALSE NEG.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
FT METAL 72 72 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 86 86 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 173 173 IRON 2 (HEME B562 AXIAL LIGAND).
FT METAL 187 187 IRON 1 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 391 AA; 46044 MW; FCF34BAC4A543C25 CRC64;

Query Match 5.18; Score 98.5; DB 1; Length 391;
Best Local Similarity 21.58; Pred. No. 0.59;
Matches 79; Conservative 58; Mismatches 131; Indels 99; Gaps 20;

Qy 25 DIVSCVGMFFLLGLVFGTAASIVFTLQHSVAVPAEQAQSGSKSLYYGVKDLATVF 84
Db 17 EVFSLFGFTFTIIVQLVSGTMLAFSSVPEPLIPTVRDEE-----DIEDLYTDD 67

Qy 85 FYMLVAIIHATIQEYVLDKINKRMQF---TKAKQNFN-----ESQFSVVFYFSCI 134
Db 68 FFWL-----HERGVDLIFISYFHLRLKLYLVNFDLETAESWKSQGVSFVLQVWV 118

Qy 135 WGTFILISENCLSDPTLIWKARP--HSMMTFQMKFFYISQLAYWFHAFPELYFQTKKODI 193
Db 119 FFGVLVLCCTH--LSEITLITIAANIFHTFEMFKGK-----AYWF-----LFTDQLNTDT 165

Qy 194 PRQLVYIGLHLFHITGAYLLYLNHLVLLVYFVLLSHMCGLFY--FSDEKYQKGLS- 251
Db 166 LIRLAYA-----HVVSAP--YLSFLGLL-----HGIDIHWDKNEPFYDGLSS 206
```

THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1357.5	70.7	374	11	Q91V04	mus musculus
2	1251	65.2	373	13	Q90ZM0	xenopus lae
3	1129	58.8	369	13	Q90ZM1	brachydanio
4	820	42.7	371	13	Q90ZL9	xenopus lae
5	811	42.2	370	11	Q92L45	mus musculus
6	602.5	31.4	368	5	Q9W5C2	Q9W5C2 drosophila
7	602.5	31.4	1575	5	Q9W5C3	Q9W5C3 drosophila
8	597.5	31.1	368	5	Q9U1L3	Q9U1L3 drosophila
9	494	25.7	371	5	Q9U3P5	Q9U3P5 caenorhabdi
10	487	25.4	373	5	Q9XXK7	Q9XXK7 caenorhabdi
11	472.5	24.6	159	11	Q9CVJ6	Q9CVJ6 mus musculus
12	156.5	9.2	393	11	Q9D5J1	Q9D5J1 mus musculus
13	146.5	7.6	308	10	Q9M6A4	lycopersico
14	143	7.4	394	4	Q9HA82	homo sapien
15	134.5	7.0	400	5	Q95RN6	Q95RN6 drosophila
16	127	6.6	310	10	Q9LDF2	Q9LDF2 arabidopsis

Db	121	ESGOLSAYFLACVWGTFILISENYISDPITILWRAPPHNLMTFQTKFFVISQIAYWLHAF	180
Qy	181	PELYFQTKKKQDIPROLVYIGLHLFHITGAYLLYLNLHGLLLLVLHYFVELLSHMGCLFY	240
Db	181	PELYFQTKKKEDIPROLVYIGLHLFHITGAYLLYLNLHGLLLLVLHYFVELLSHMGCLFY	240
Qy	241	FSDEKYQKGISLWAIIVFILLGRVTLIVSVLTVGFFHLAGSQNRNPDALTCNVNVLAAKIAV	300
Db	241	FSDEKYQKGSFSLWAVFLVGRLLTLILSVLTVGFGIARAENOKLDSTGNFNVLAVRIAV	300
Qy	301	LSSSCTTQAVYVWNILPLTWLQWRVWDSNIQASCMKKK---RSRSKKRKTENG-VGVTS	355
Db	301	LASICITQAEWMKMFNFQLRRRREHSAFOAPPVKRKPATVKGRRSSKRGKTENGVGNTVTS	360
Qy	356	NRVDCPPKRKEKSS	369
Db	361	NGADSPNRKKEKSS	374
RESULT	2		
Q90ZM0			
ID	Q90ZM0	PRELIMINARY;	PRT; 373 AA.
AC	Q90ZM0;		
DT	01-DEC-2001	(TEMBLrel. 19, Created)	
DT	01-DEC-2001	(TEMBLrel. 19, Last sequence update)	
DT	01-DEC-2001	(TEMBLrel. 19, Last annotation update)	
DE	TRAM1		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipioidea; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI_TaxID=8355;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RL	Hartmann E.;		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY029534; AAK40302.1; -.		
SQ	SEQUENCE 373 AA; 43121 MW; 62AP49E9859769DF CRC64;		

Query Match 42.7%; Score 820; DB 13; Length 371;
 Best Local Similarity 44.3%; Pred. No. 4.5e-67;
 Matches 170; Conservative 73; Mismatches 111; Indels 30; Gaps 10;

QY 1 MGLRKKSTKNPPVLSQBEFILONHADIVSCVGMFF-----LLGLVPFGTAEASIVFTLQHS 56
 DB 1 MAFRRR--KSYPLFSQEFVIINHADI-----GFFVFLCVLIGLMEFVTAFTAFITLPQYN 54
 QY 57 VAVPAAEQATGSKSLYYGVKDLATVFYMLVAIIIIHATIOEYVLDKINKRMQFTKRAKQ 116
 DB 55 SSIQTLD---GETLYHYGVKDLVTILFYVYVIAIIILHAIYQVEYILDINKRHLKSVKQ 110
 QY 117 NKFNESQGSVFYFFCSIGWTFFILISENCLSDPTLIWKARPHSMMTFQMKFFYISQLAYW 176
 DB 111 SRFNSGQAAAFHLASMFCLVVSATGELYSPKTLWESYPHVYLPFQVKFFYLQCLAYW 170
 QY 177 FHAPPELYFQTKKQDIPROLVYIGLHLFHTGAYLLYLNHLGGLLLVHLHFVVELLSHMC 236
 DB 171 LHALPELYFQVKKEEVPRLQYIVLYLHLAGAYLLNLTRGLILLILQSVAEPLFHIA 230
 QY 237 GLFFVDSKQYKGTSLWAIVFIILGRVLTLYSVLTGVGFLAGSO--NRNPDALTCNVNVL 294
 DB 231 RLFTFTDENQRLNAGVGVFIITRLFTTLUSVTITGFLARAEVHTNPDP--KGTLTNL 288
 QY 295 AAKIAYLSSSCTIOAYVTWNLTITLQWRVEDSNIAQS-----CMKKRSRSKSKRT- 346
 DB 289 LFRMVVLLLMCVSQTMMWRFIHFQLRRWRRECCKEQAARKRSVAVAMKQAKYIKRESG 348
 QY 347 --ENGVGVEYSNRVDCPPKRKEKS 368
 DB 349 YHENG-VKAENG-STPRQKKIKS 370

RESULT 5
 Q92425 PRELIMINARY; PRT; 370 AA.
 AC Q92425;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE TRAM2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RN RNP SEQUENCE FROM N.A.
 RA Hartmann E.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
 DR EMBL: AY029530; AAK40298.1; -
 SQ SEQUENCE 370 AA; 43182 MW; F8E768AFB0582548 CRC64;

Query Match 42.2%; Score 811; DB 11; Length 370;
 Best Local Similarity 42.7%; Pred. No. 3e-66;
 Matches 159; Conservative 80; Mismatches 119; Indels 14; Gaps 5;

QY 1 MGLRKKSTKNPPVLSQBEFILONHADIVSCVGMFFLLGLVPFGTAEASIVFTLQHSVAP 60
 DB 1 MAFRRR-TKSYPLFSQBEFIINHADIIGFCLVLCVLIGLMEFVTAFTAFITLPQYNISVP 59
 QY 61 AAEEQATGSKSLYYGVKDLATVFYMLVAIIIIHATIOEYVLDKINKRMQFTKAKONFN 120
 DB 60 TADSETVH----YHNGPKDLVTILFYVVTITFRAVQVEYILDKISKRHLHLSKVKHSFN 115
 QY 121 ESGQFSVFYFFCSIGWTFFILISENCLSDPTLIWKARPHSMMTFQMKFFYISQLAYWHAF 180
 DB 116 ESGQLLVPHLSAVAMCFVIVTEGYLNPRLSLEDYDPHYLSFQVKFFYLQCLAYWLHSL 175
 QY 181 PELYFQTKKQDIPROLVYIGLHLFHTGAYLLYLNHLGGLLLVHLHFVVELLSHMCGLFY 240


```
RP SEQUENCE FROM N.A.
RA Papagiannakis G., Spanos L., Bolshakov V., Siden-Kiamos I., Louis C.;
RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL109630; CAB65875.1; -;
DR FlyBase: FBgn0040340; EG:BACR7A4.5.
SQ SEQUENCE 368 AA; 41766 MW; 693795FFFC2ED4A6 CRC64;

Query Match 31.1%; Score 597.5; DB 5; Length 368;
Best Local Similarity 36.6%; Pred. No. 1.1e-46;
Matches 137; Conservative 63; Mismatches 133; Indels 41; Gaps 8;

QY 4 RKSTKNPVLQOEFLQNHADIVSCVGMFFLLGLVFEGETAEASIVFLTLQHSVA-VPAA 62
DB 9 RKTSTKNPVLQOEFLQNHADIVSCVGMFFLLGLVFEGETAEASIVFLTLQHSVA-VPAA 62
QY 63 EQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKQNFES 122
DB 69 RQPYGKPTYYTAGIKDYCAIFFYTLTCIIMHAIQEFVLDNISKKLHLSKFLARFES 128
QY 123 GQSVYFFSCIWGTFFILISENCLSDPTLIWKARPHSMFTFQMKFFYISQLAYWPHAFPE 182
DB 129 GOLVAFYLLSFVNGAHVLLKEGVLGOVQWEGFDPHMSFLHKFFYVQVLAAYLHMLPE 188
QY 183 LVFQ--KTKKQDIPROLYI-GLHFLHITGAYLLYLNHGLLGLVFEVLLSHMCGLGF 239
DB 189 LVFQKIKTEEQPKIVHSISGFTL--IVLAYTSLQRLALVLLTLHYFSELLSHVFLI 246
QY 240 --YFDEKQKGLSLWAIYVILGLRLVTLIVSVLTGVLGFLAGSQNRNPDALTGNVNLAK 297
DB 247 GVFDREERLAKLRVNNVAFVFLIRFATSVIGVTLTYIGGVRS----- 290
QY 298 IAVLSSSCVITQAVYVWNLTITLWLRQWVEDSNIOASCMKKRSKKRTENGVGVSNSR 357
DB 291 LLALGGLIALQGLYVFSFT-----EQLRAKREAKREAKLALQTKKP 336
QY 358 VDCPP---RKREKS 368
DB 337 AKTPKDKVRKKES 350

RESULT 9
QY03P5 PRELIMINARY; PRT; 371 AA.
AC Q903P5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE C24F3.1A PROTEIN.
GN C24F3.1A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: AL022716; CAA18770.1; -.
DR InterPro: IPR001185; MSCL.
SQ SEQUENCE 371 AA; 42420 MW; 5269FC7231222592 CRC64;
```

```
Query Match 25.7%; Score 494; DB 5; Length 371;
Best Local Similarity 32.4%; Pred. No. 3.5e-37;
Matches 119; Conservative 74; Mismatches 144; Indels 30; Gaps 9;

QY 2 GLRKKSTKNPPVLSQEFILQNHADIVSCVGMFFLLGLVFEGETAEASIVFLTLQ----HSV 57
DB 7 GSKASKKPPPLSLSHFEIIONHGDIMSCVVMVFIIVGLMPLTHLSLSLFIAPQYNGTYTV 66
QY 58 AVPAEEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKQN 117
DB 67 AVEQGOEREVHG---YLSGLDLPALFAFFYSVCWIVHVVAVVQYEGLDKISKTHLSKVSTF 123
QY 118 KNESQSFVFFFCIWGTFFILISENCLSDPTLIWKARP--HSMMTFQMKFFVLSQL 173
DB 124 KGESFQHMFTVYS-IAHAFYIVSERLEDFSEVSKVGLGYPFTEHRVMSAAYKLYIFQI 182
QY 174 AYWFHAFPELYFOKTKKQDIPROLYI-GLHFLHITGAYLLYLNHGLLGLVFEVLLS 233
DB 183 SYWIHQFPFYLQKLRDEIROKSVQAILHIAFISTAYFNFTRVGLALITLEYITQLIF 242
QY 234 HMCGLYFDEKVKIGIS-----LMAIVILGLRLVTLIVSVLTGVLGFLAGSQNRNPDALT 288
DB 243 HIARFAHFGV---RKLSDPAFKLNGSFVLRLSGIIIAVMTFWYGLRQAESPFVDISA 299
QY 289 GNVNLAATAVLSSCTIQAVYVWNLTITLWLRQWVEDSNIOASCMKKRS-----RS 341
DB 300 GNFNTAVIRLNVLLAVVLLQLFLLYSFVFMHGMRFRESN---AKKEKKSAAMAAVPRK 356
QY 342 SKKRTEN 348
DB 357 EKKRQDS 363

RESULT 10
QY0XK7 PRELIMINARY; PRT; 373 AA.
AC Q9XK7;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE C24F3.1B PROTEIN.
GN C24F3.1B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: AL022716; CAA18772.1; -.
DR InterPro: IPR001185; MSCL.
SQ SEQUENCE 373 AA; 42655 MW; 92D65DF05F50B9CC CRC64;

Query Match 25.4%; Score 487; DB 5; Length 373;
Best Local Similarity 32.4%; Pred. No. 1.5e-36;
Matches 120; Conservative 73; Mismatches 143; Indels 34; Gaps 10;

QY 2 GLRKKSTKNPPVLSQEFILQNHADIVSCVGMFFLLGLVFEGETAEASIVFLTLQ----HSV 57
DB 7 GSKASKKPPPLSLSHFEIIONHGDIMSCVVMVFIIVGLMPLTHLSLSLFIAPQYNGTYTV 66
QY 58 AVPAEEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKQN 117
```



```
QY 167 FFYISQLAYWFHAPPELYFOKTKQDIPROLVY-----IGLHLFHTGAYLLYLNLHGLLL 222
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 WYLLLEGLFVLSLLTLPFD-VKRDKFEQVHHFVAVGL-----IGFSYSWNLLRIGAVV 235
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 223 LVLHYFVELLSHMCGLFYFSDKEQKQKIS-----LWAIIVFILGRVLTVLIVSVL 270
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 236 LLLHDCSDYLLEGCKILNYA--HFRRGCDALFIMFALVFFYTRLIFFPTQVI 285
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
Q9M6A4 ID Q9M6A4 PRELIMINARY; PRT; 308 AA.
AC Q9M6A4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE ASCL.
GN ASC.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. VENT CHERRY;
RX MEDLINE=99168767; PubMed=10071209;
RA Mesbah L.A., Kneppers T.J., Takken F.L., Laurent P., Hille J.,
RA Nijkamp H.J.;
RT "Genetic and physical analysis of a YAC contig spanning the fungal
RT disease resistance locus Asc of tomato (Lycopersicon esculentum).";
RL Mol. Gen. Genet. 261:50-57(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. VENT CHERRY;
RX MEDLINE=20243803; PubMed=10781105;
RA Brandwagt B.F., Mesbah L.A., Takken F.L.W., Laurent P.L.,
RA Kneppers T.J.A., Hille J., Nijkamp H.J.J.;
RT "A longevity assurance gene homolog of tomato mediates resistance to
RT Alternaria alternata f. sp. lycopersici toxins and fumonisins B1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:4961-4966(2000).
DR EMBL; AF198177; AAF67518.1; -;
SQ SEQUENCE 308 AA; 36301 MW; 2BD3378CE53F416D CRC64;

Query Match 7.68; Score 146.5; DB 10; Length 308;
Best Local Similarity 23.18; Pred. No. 2e-05;
Matches 50; Conservative 47; Mismatches 78; Indels 41; Gaps 9;

QY 78 KDLATVFYMLVAIIHATIQEYVLDKINKRMQFTK-----AKONKFNESQGF 125
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 22 QDLFLFFLFFVFLRFLDRFVEALAKRMIFGKTVVNVINGREERKKINKKESAWK 81
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 126 SVFYFFSCINGTFLISENCLSDPTLIWKAR-----PHSMMTFQMKFFYISQLAYWFHA- 179
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 82 FVYFSLAELLALSVCNEPWFDSRYFAGDGVVWPNLKMKLKLLYNYAGGFYFSI 141
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 180 PPELYFOKTKQDIPROLVYIGLHLFHTGAYLLYLNH-----LGLLLVLH-----YFV 229
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 142 FATLYWE--TRYDFAAQIIH-----HVTTVSLVLSVYGFARIGSVLALHDGSDVFM 194
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 230 EL--LSHMCGLFYFSDKEQKQKISLWAIIVFILGRVL 263
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 195 EIAKMSKSYGPDLIADIFF-----SLFALVFTSLRII 226
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
Q9HA82 ID Q9HA82 PRELIMINARY; PRT; 394 AA.
AC Q9HA82;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
```

```
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 46.4 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwavanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG CARCINOMA;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK022151; BAB13972.1; -;
DR EMBL; BC009828; AAH09828.1; -;
DR InterPro; IPR001356; Homeobox.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 394 AA; 46398 MW; 67228829BDED2801 CRC64;

Query Match 7.48; Score 143; DB 4; Length 394;
Best Local Similarity 25.48; Pred. No. 5.6e-05;
Matches 46; Conservative 39; Mismatches 84; Indels 12; Gaps 5;

QY 108 RMOFTKAKQKNFSGQFVFFSCIWGTFILISENCLSDPTLIWKARPHSMMTFQMKF 167
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 126 RPQLTK-----KFCASWRFLFSLSFVGLSVLYHESLVWAPVWCDRYPNQTLKPSLYW 181
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 168 FYISQLAYWFHAPPELYFOKTKQDIPROLVYIGLHLFHTGAYLLYLNLHGLLLVLHY 227
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 182 WYLLLEGLFVLSLLRPLFD-VKRDKFEQVTHHFVAVILMTFTFSYANLLRIGSLVLLHD 240
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 228 FVELLSHMCGLFYFSDKEQKQKIS-----LWAIIVFILGRVLTVLIVSVLTVGFHAGSONRN 283
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 SSDLLEACKMVNYM--QYQVCDALFLIFSFFYFTRLVFLFPTQLITYTESIS-NRG 297
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 284 P 284
      |
Db 298 P 298

RESULT 15
Q95RN6 ID Q95RN6 PRELIMINARY; PRT; 400 AA.
AC Q95RN6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DR DLI8904P.
GN CG15898.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=y, CN BW SP;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
```

[illegible]

Search completed: September 6, 2002, 17:20:27
Job time: 636 sec